Biomarkers and risk factors of neurodegeneration in multiple sclerosis

Ingrid Anne Lie

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2023

UNIVERSITY OF BERGEN

Biomarkers and risk factors of neurodegeneration in multiple sclerosis

Ingrid Anne Lie

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 10.02.2023

Year: 2023

Title: Biomarkers and risk factors of neurodegeneration in multiple sclerosis

Name: Ingrid Anne Lie

Print: Skipnes Kommunikasjon / University of Bergen

Scientific environment

This thesis was performed at the Department of Clinical Medicine, University of Bergen, Norway, the Structural Brain Imaging Group, Department of Radiology and Nuclear Medicine, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, location VUmc, Amsterdam, The Netherlands and Neuro-SysMed, Department of Neurology, Haukeland University Hospital, Bergen, Norway.

Main supervisor: Professor Lars Bø

The Norwegian Multiple Sclerosis Competence Centre, Department of Neurology, Haukeland University Hospital, Norway.
Department of Clinical Medicine, University of Bergen, Norway.

Co-supervisors: Professor Kjell-Morten Myhr

Neuro-SysMed, Department of Neurology, Haukeland University Hospital, Bergen, Norway.
Department of Clinical Medicine, University of Bergen, Norway.

Professor Øivind Torkildsen

Department of Clinical Medicine, University of Bergen, Norway.
Neuro-SysMed, Department of Neurology, Haukeland University Hospital, Bergen, Norway.

Associate Professor Hugo Vrenken

Department of Radiology and Nuclear Medicine, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, location VUmc, Amsterdam, The Netherlands.

Funded by: University of Bergen

Acknowledgements

Most of all I would like to express my gratitude to my supervisors. To my main supervisor, Professor Lars Bø, thank you for the trust, encouragement, and support to make this project my own. You are a great inspiration, of course as an experienced neurologist and accomplished researcher, but also as a kind and considerate person. To my co-supervisor, Professor Kjell-Morten Myhr, thank you for your generous sharing of resources, ideas, research know-how and overall enthusiasm! I deeply admire your drive and dedication. For invaluable guidance since the first day of my neurological residency, and for quickly (and perhaps consciously) sparking my interest in multiple sclerosis, I wish to thank my co-supervisor, Professor Øivind Torkildsen. You are such an educator, and I feel lucky I got to continue learning from you. Lastly, this project would not be the same without my co-supervisor, Associate Professor Hugo Vrenken. Thank you for finding a place for me in your research group, and for expertly leading the way through all the challenging, unfamiliar, frustrating, rewarding, and amazing parts of a PhD project.

I sincerely thank the people participating in the OFAMS study and the 10-year follow-up study, the OFAMS study group and all personnel who have contributed to these projects. For invaluable input on my four papers, thank you to my co-authors Frederik Barkhof, Iman Brouwer, Alla Bru, Astrid Edland, Randi Eikeland, Sonia Gosal, Hanne F. Harbo, Trygve Holmøy, Sezgi Kacar, Emma Kerklingh, Grethe Kleveland, Silje S. Kvistad, Rozemarin M. Mattiesing, Rune Midgard, Marijke A.E. Mol, David R. van Nederpelt, Petra J.W. Pouwels, Yvonne S. Sørenes, Charlotte E. Teunissen, Kristin N. Varhaug, Christian Vedeler, Merlin Weeda, Stig Wergeland, Kristin Wesnes and Nina Øksendal.

Next, I want to acknowledge the talented and skillful members (and former members) of the Structural Brain Imaging Group (SBIG): Georgios Lappas, David van Nederpelt, Samantha Noteboom, Ronald van Schijndel, Alexandra de Sitter, Merlin Weeda, Alle Meije Wink and Viktor Wottschel. A very special thanks to Iman Brouwer, for all the

hours of working (magic) on the OFAMS data. I am forever grateful. To Rozemarijn M. Mattiesing, thank you for your friendship both inside and outside the office.

As part of the MS-research group at Haukeland University Hospital, I have enjoyed the company of the many experienced professionals and lovely people belonging to this environment. I am particularly thankful for my fellow PhD students, travel buddies, Rotunden neighbors, and friends: Karine Eid, Hilde Norborg, Brit Ellen Rød and Hilde Marie Torgauten.

Through the years of this project, I have been cheered on by my fun, warm and supportive friends. Some of whom I have been lucky to gain along the way, and some I am grateful to have had from the beginning. I especially want to thank Nina, my friend in life, for all the joy and care.

Finally, to my parents, Betty and Olav, thank you for being the foundation and support I always know I have. Dear brothers, Torstein and Kjetil, I guess an academic paper is a fitting place for me to express my appreciation for you two. I truly value the time I spend with all of you, and everything we get to experience together as a family.

Bergen, September 2022

Ingrid Anne Lie

Abbreviations

9-HPT	9-hole peg test
BBB	Blood-brain barrier
BMI	Body mass index
CIS	Clinically isolated syndrome
CNS	Central nervous system
CSF	Cerebrospinal fluid
DMT	Disease modifying therapy
EDSS	Expanded Disability Status Scale
ELISA	Enzyme-linked immunosorbent assay
FFE	Fast field echo
FLAIR	Fluid-attenuated inversion recovery
FLASH	Fast low angle shot
FLIRT	FMRIB`s Linear Image Registration Tool
FSL	FMRIB Software Library
GBCA	Gadolinium-based contrast-agent
Gd+	Gadolinium-enhancing
GFAP	Glial fibrillary acidic protein
GM	Grey matter
ICC	Intraclass correlation coefficient
LC	Lesion count
LST	Lesion Segmentation Tool
OFAMS	ω-3 fatty acids in multiple sclerosis
MPRAGE	Magnetization-prepared rapid acquisition gradient echo
MRI	Magnetic resonance imaging

MS	Multiple sclerosis
NEDA	No evidence of disease activity
NEX	Number of excitations
NfH	Neurofilament heavy
NfL	Neurofilament light
NfM	Neurofilament middle
NO	Nitric oxide
PASAT	Paced Auditory Serial Addition Test
PP	Primary progressive
RCT	Randomised controlled trial
PNS	Peripheral nervous system
RR	Relapsing-remitting
sNfL	Serum neurofilament light
SDMT	Symbol Digit Modalities Test
SIMOA	Single-molecule array
SP	Secondary progressive
T	Tesla
T25FW	Timed 25-Foot Walk
TE	Echo time
TI	Inversion time
TR	Repetition time
WM	White matter

Abstract in English

Background and objectives: Markers of neurodegeneration are closely related to clinical disability and disease progression in multiple sclerosis (MS). Hence, neurodegenerative biomarkers may help identify future therapeutic targets, measure the effect of initiated treatment against such targets, and clarify how environmental factors and comorbid conditions may affect neurodegenerative processes.

Clinical implementation of candidate biomarkers is challenged by several factors. First, the neurodegenerative pathological substrate is unclear, especially to which degree neurodegenerative changes occur secondary to inflammatory damage in the white matter (WM), or due to processes affecting the grey matter (GM) primarily. Second, variability in measurements related to technical, physiological and disease related variations, comorbid conditions and environmental factors needs clarification. The main objective of this study was to address some of these challenges. In more detail, we sought to 1) explore the spatio-temporal relationship between WM lesions and global and regional GM atrophy in patients with MS, and if the association differs in the clinical phenotypes, 2) investigate whether reliable brain atrophy measurements can be obtained from 3D T1-weighted images acquired after administration of gadolinium-based contrast agents (GBCAs), using FreeSurfer, 3) investigate how serum neurofilament light (sNfL) levels measured during, and outside of periods of inflammatory disease activity, associate with GM atrophy and clinical disability ten years later in patients with relapsing-remitting MS (RRMS) and 4) assess whether smoking in patients with RRMS relate to GM atrophy, lesion load and clinical disability after ten years.

Methods: In article I, we performed a systematic review including qualitative and descriptive analyses. MEDLINE and Embase were searched for abstracts containing direct associations between brain GM and WM lesion measures obtained by conventional MRI sequences in patients with clinically isolated syndrome (CIS) and MS. In articles II, III and IV, we used data from a multicentre, randomised trial of ω-3 fatty acids (the OFAMS study) in people with RRMS, and from the follow-up visit

conducted approximately ten years after the conclusion of the OFAMS study. Patients underwent clinical, biochemical and MRI examinations regularly during the 24 months of the OFAMS study, and once at the 10-year follow-up visit. In article II, 22 patients were included, in which 3D T1-weighted MR images were obtained during the same scanner visit, both before and after administration of GBCAs. The difference between measurements obtained in pre- and post-contrast images was assessed by paired t-tests, and the consistency by intra-class correlation coefficient (ICC). In articles III and IV, 78 and 85 patients were included, respectively. The association between sNfL levels and smoking and long-term outcome measures was investigated by linear multilevel or regular regression models. All models were adjusted for age and sex, in addition to confounders relevant to specific analyses.

Results: In the first article, 90 studies were included. Higher WM lesion load was in the majority of studies associated to more GM atrophy, with the most consistent relationship found in in early (relapsing) disease, and less so in progressive MS.

In the second article, we found good to excellent consistency between all values obtained from pre- and post-contrast images (ICC ranging from 0.926 to 0.996). In post-contrast images, cortical thickness and GM volumes were significantly higher than in pre-contrast images, while total WM volume was significantly lower.

In article III, higher mean sNfL levels during periods with gadolinium-enhancing lesions present or recently present (representing inflammatory disease activity) predicted lower mean cortical thickness, lower total and deep GM volume (standardised β ranging from -0.399 to -0.581) and higher T2 lesion count (standardised β=0.498) ten years later. Higher inflammatory sNfL levels were also associated with higher score (higher disability) on the dominant hand 9-hole peg test (standardised β=0.593). No MRI or clinical outcome measures were associated with higher sNfL levels during periods with no gadolinium-enhancing lesions present or recently present (remission).

In article IV, smoking in patients with RRMS was associated with lower total WM and deep GM volume, and higher T2 lesion volume after ten years. Of the clinical

outcomes, smoking was associated with higher walking impairment measured by the Timed 25-Foot Walk test, and a larger decrease in Paced Auditory Serial Addition Test (attention) scores.

Conclusion: The findings in article I suggest the overall GM neurodegeneration during early disease stages may largely be secondary to damage in the WM, while in progressive MS, neurodegeneration becomes more detached, possibly dominated by primary disease mechanisms. In article II, we found that reliable brain atrophy measurements may be possible to extract from post-contrast T1-weighted images, using FreeSurfer. This finding may allow a considerable amount of historical and prospective real-world data to be used to measure brain atrophy in patients with MS. The notion of a strong relationship between inflammatory WM damage and GM atrophy is supported by the findings in article III, where sNfL levels during inflammatory disease activity, but not during remission, predicted future GM atrophy in patients with RRMS. The additional association with higher long-term disability suggests that inflammatory sNfL levels may be used to quantify the extent of ongoing axonal damage, indicating the risk of disability accrual. In article IV, smoking was associated with lower brain volumes, higher lesion load and higher disability after ten years. Whether the association with GM atrophy is caused by a direct neurotoxic effect, or by secondary neurodegeneration following heightened inflammatory activity, needs to be clarified in future studies. Nevertheless, our findings are in support of smoking patients with MS being routinely offered advice and aid in smoking cessation.

Abstract in Norwegian

Bakgrunn og mål: Markører på nevrodegenerasjon er tett knyttet til klinisk funksjonsnedsettelse og sykdomsprogresjon hos pasienter med multippel sklerose (MS). Slike biomarkører kan derfor bidra til å identifisere potensielle terapeutiske mål, måle effekten av behandling rettet mot disse, og belyse hvordan miljø- og livsstilsfaktorer og komorbide tilstander kan påvirke nevrodegenerative prosesser.

Implementering av kandidatmarkører i klinisk praksis utfordres av flere faktorer. For det første er de patologiske prosessene som fører til nevrodegenerasjon ikke tilstrekkelig kartlagt, spesielt i hvilken grad nevrodegenerative forandringer oppstår sekundært til inflammatorisk skade i hvit substans, eller som følge av prosesser som rammer grå substans primært. For det andre må variabilitet i målinger som følge av tekniske, fysiologiske og sykdomsrelaterte variasjoner, komorbide tilstander og miljø- og livsstilsfaktorer avklares. Hovedmålet med denne studien var å møte noen av disse utfordringene. Vi ønsket å 1) utforske det spatiotemporale forholdet mellom hvitsubstanslesjoner og global og regional atrofi i grå substans hos pasienter med MS, og om forholdet endret seg blant de kliniske fenotypene, 2) undersøke om 3D T1-vektede MR-bilder tatt etter administrering av gadoliniumholdige kontrastmidler kan brukes til å måle hjerneatrofi, ved bruk av FreeSurfer, 3) undersøke hvordan nivåer av nevrofilament lettkjede i serum (sNfL) målt i perioder med og uten inflammatorisk sykdomsaktivitet er assosiert med atrofi i grå substans og klinisk funksjonsnedsettelse etter ti år, hos personer med attakkpreget MS (RRMS) og 4) undersøke om røyking hos personer med RRMS er assosiert med atrofi i grå substans, lesjonsbyrde og klinisk funksjonsnedsettelse etter ti år.

Metode: I artikkel I utførte vi en systematisk oversikt over den eksisterende litteraturen, samt kvalitative og deskriptive analyser. Vi søkte i MEDLINE og Embase etter rapporter som inneholdt direkte assosiasjoner mellom mål på atrofi i grå substans og lesjoner i hvit substans i hjernen. Målene måtte være hentet fra konvensjonelle MR-bildesekvenser, hos pasienter med klinisk isolert syndrom (CIS) og MS. I artikkel II, III og IV brukte vi data fra multisenterstudien «ω-3 fatty acids in MS» (OFAMS-

studien), og fra oppfølgingsbesøket gjennomført omtrent ti år etter OFAMS-studien ble avsluttet. Totalt 92 pasienter med RRMS deltok i OFAMS-studien, og gjennomgikk jevnlige kliniske, biokjemiske og MR-undersøkelser i løpet av de 24 månedene studien pågikk. Åttifem av pasientene deltok på 10-års oppfølgingsbesøket, hvor undersøkelsene ble gjentatt. Totalt 22 pasienter, hvor 3D T1-vektede MR-bilder ble tatt både før og etter administrasjon av gadoliniumholdig kontrastmiddel ble inkludert i artikkel II. Intraklassekoeffisient (ICC) ble benyttet i reliabilitetsanalysen, og parede t-tester ble benyttet for å vurdere eventuelle forskjeller i målene. Totalt 78 og 85 pasienter ble inkludert i henholdsvis artikkel III og IV. Effekten av sNfL nivå og røyking på langsiktig hjerneatrofi og klinisk funksjonsnedsettelse ble undersøkt ved blandede og konvensjonelle lineære regresjonsmodeller. Alle modellene ble korrigert for alder og kjønn, i tillegg til eventuelle konfundere relevante for spesifikke analyser.

Resultat: Totalt 90 studier ble inkludert i artikkel I. Høyere lesjonsbyrde i hvit substans var i de fleste studiene assosiert med mer atrofi i grå substans. Forholdet var sterkest og hyppigst sett ved tidlig (attakkpreget) sykdom, og i mindre grad ved progressiv MS.

I artikkel II fant vi utmerket konsistens mellom alle verdier hentet fra MR-bilder tatt før og etter administrasjon av kontrastmiddel (ICC verdier fra 0.926 til 0.996). Ved sammenlikning av målinger hentet fra bilder tatt før og etter administrasjon av kontrastmiddel, var kortikal tykkelse og grå substans volum på bilder tatt etter kontrastmiddel signifikant høyere enn på bilder tatt før kontrastmiddel, mens det totale volumet av hvit substans var signifikant lavere.

I den tredje artikkelen fant vi at høyere gjennomsnittsnivå av sNfL i perioder med aktiv inflammasjon (kontrastladende lesjoner til stede eller nylig til stede) predikerte lavere volum av grå substans i hele hjernen og dyp grå substans, lavere kortikal tykkelse (standardisert β fra -0.399 til -0.581) og høyere antall T2 lesjoner i hvit substans (standardisert β=0.498) etter ti år. Høyere nivå av inflammatorisk sNfL var også assosiert med større funksjonsnedsettelse (høyere skår) målt ved dominant hånd 9-hole peg test (standardisert β=0.593). Ingen MR- eller kliniske mål var assosiert med høyere

sNfL nivå i perioder med remisjon (ingen kontrastladende lesjoner til stede eller nylig til stede).

I artikkel IV fant vi at røyking hos pasienter med RRMS var assosiert med lavere totalt volum av hvit substans og dyp grå substans, og høyere volum av T2 lesjoner etter ti år. Av de kliniske målene var røyking assosiert med nedsatt gangfunksjon målt ved Timed 25-Foot Walk test, og større nedgang i oppmerksomhetsskår, målt ved Paced Auditory Serial Addition Test.

Konklusjon: Funnene i artikkel I tyder på at nevrodegenerasjon i grå substans i stor grad oppstår sekundært til skade i hvit substans ved tidlige (attakkpregede) sykdomsstadier av MS. Ved progressiv sykdom derimot, virker det å være en gradvis overgang hvor nevrodegenerasjonen domineres av sykdomsmekanismer som rammer grå substans primært. I artikkel II fant vi at pålitelige mål på hjerneatrofi kan hentes fra T1-vektede bilder tatt etter administrasjon av kontrastmiddel, ved bruk av FreeSurfer. Dette funnet kan åpne for bruk av en betydelig mengde historiske og prospektive real-world data til å måle hjerneatrofi hos pasienter med MS. I artikkel III fant vi, i tråd med resultatene fra artikkel I, at sNfL nivå i perioder med aktiv inflammasjon, men ikke remisjon, predikerte fremtidig atrofi i grå substans hos pasienter med RRMS. Inflammatorisk sNfL nivå var også assosiert med langsiktig funksjonsnedsettelse. Dette tyder på at målingene kan brukes til å anslå omfanget av pågående aksonal skade under et attakk, og indikere risiko for varig nedsatt funksjonsevne. I artikkel IV fant vi at også røyking predikerte økt hjerneatrofi, høyere lesjonsbyrde og økt funksjonsnedsettelse etter ti år. Funnene taler for at pasienter med MS som røyker bør rutinemessig tilbys rådgivning og hjelp til røykeslutt. Om assosiasjonen med atrofi i grå substans skyldes en direkte nevrotoksisk effekt av røyking, eller en sekundær nevrodegenerasjon via økt inflammatorisk aktivitet, bør undersøkes nærmere i fremtidige studier.

List of Publications

I. 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(15):e1562-e1573.

II. Lie IA, Kerklingh E, Wesnes K, van Nederpelt DR, Brouwer I, Torkildsen Ø, Myhr KM, Barkhof F, Bø L, Vrenken H. The effect of gadolinium-based contrast-agents on automated brain atrophy measurements by FreeSurfer in patients with multiple sclerosis. *Eur Radiol.* 2022;32(5):3576-3587.

III. Lie IA, Kaçar S, Wesnes K, Brouwer I, Kvistad SS, Wergeland S, Holmøy T, Midgard R, Bru A, Edland A, Eikeland R, Gosal S, Harbo HF, Kleveland G, Sørenes YS, Øksendal N, Varhaug KN, Vedeler CA, Barkhof F, Teunissen CE, Bø L, Torkildsen Ø, Myhr KM, Vrenken H. Serum neurofilament as a predictor of 10-year grey matter atrophy and clinical disability in multiple sclerosis: a longitudinal study. *J Neurol Neurosurg Psychiatry.* 2022;93:849-857.

IV. Lie IA, Wesnes K, Kvistad SS, Brouwer I, Wergeland S, Holmøy T, Midgard R, Bru A, Edland A, Eikeland R, Gosal S, Harbo HF, Kleveland G, Sørenes YS, Øksendal N, Barkhof F, Vrenken H, Myhr KM, Bø L, Torkildsen Ø. The effect of smoking on long-term gray matter atrophy and clinical disability in patients with relapsing-remitting multiple sclerosis. *Neurol Neuroimmunol Neuroinflamm.* 2022;9(5)e200008.

Contents

1. Introduction

Multiple sclerosis (MS) is one of the most common chronic neurological disorders. Globally, around 2.8 million people live with MS, with the highest prevalence in Europe and America.[1] The disease typically debuts in early adulthood and is more than twice as likely to develop in females.[1] MS carries a great personal and societal cost. Physical disability, cognitive difficulties and fatigue may be detrimental to the work and social life of people with MS, while health-care costs and loss of production burden the society.[2]

There is currently no curative treatment available. Developments in disease modifying therapies (DMTs) targeting immunological and inflammatory disease mechanisms have significantly improved the prognosis in some groups of patients.[3] However, therapies specifically against neurodegenerative processes are still missing, and will be crucial to be able to successfully treat patients in all phases of the disease. In the process of developing neuroprotective treatments, biomarkers reflective of neurodegenerative disease mechanisms are needed throughout: from identifying possible therapeutic targets, to measuring the effects of initiated treatments, environmental factors and comorbid conditions.

1.1 Pathogenesis and aetiology of multiple sclerosis

1.1.1 Pathogenesis

The initial event or pathological trigger causing MS has not been identified. Currently, two opposing theories are most commonly considered: the most widely recognized hypothesis describes MS as primarily an autoimmune disease, starting with autoreactive T cells recognizing self-epitopes in the central nervous system (CNS) as peptide epitopes derived from pathogens.[4] After being activated in the peripheral tissues, the T cells migrate to the lymph nodes, initiating a proinflammatory cascade, in which antigen-specific T cells, B cells and plasma cells (differentiated B cells), invade the CNS.[5] Here, plasma cells release antibodies targeting both the myelin sheath and glial cells, while cytokines and other inflammatory mediators released by the

activated lymphocytes open the blood-brain barrier (BBB), attracting additional monocytes and lymphocytes.[5,6] These inflammatory and demyelinating processes are classically limited to focal regions, known as MS lesions. Hence, the MS lesion formation may first be initiated by an aberrant adaptive immune response, then largely mediated by innate immunity and activated phagocytes,[6] in a pristine CNS.

Opposed to the suggested "outside-in" hypothesis, in which the pathologic trigger initiates outside of the CNS, an "inside-out" model has gained attention in recent years. Proposing the initiating event to occur *within* the CNS, possibly as a primary defect or dysfunction in oligodendrocytes and myelin,[7] some researchers have characterised MS as primarily a degenerative disease.[8] Following the primary cytodegenerative process, a release of antigenic debris is thought to trigger a secondary aberrant autoimmune response, in genetically and environmentally predisposed individuals.[8]

1.1.2 Aetiology

MS is in part caused by the effect of over 200 gene regions associated with an increased disease susceptibility, each providing a minor effect. The identified gene regions are involved with immune mechanisms, of both the adaptive and innate immune system.[9]

While there is an undisputable link between genetic predisposition and development of MS,[9,10] it does not fully account for the individual disease risk, highlighting the role of environmental exposures. Increased risk of MS is associated with several environmental factors, of which Epstein-Barr virus infection is the most important, with a 32-fold increase in risk of MS after infection.[11] Other known environmental risk factors are low levels of vitamin D,[12] childhood-[13] and adult[14] obesity and smoking. Ever-smokers are shown to have a higher risk of MS compared to never-smokers, in a dose-dependent manner.[15,16] Furthermore, the risk is higher in current smokers than in past smokers and never-smokers, and higher in past smokers than in never-smokers.[17]

1.2 Pathophysiology

1.2.1 Focal inflammation

In 1868, Jean Martin Charcot (1825-93), a French neurologist, described a disease of the nervous system that he named "Sclérose en plaques",[18] later termed "multiple sclerosis" in the English literature. His identification of MS as a single nosological entity was based on clinical observation paired with the detailed description of the main pathological characteristic of the disease: the MS plaque, or the MS lesion.[19]

Since then, great efforts have been made to further describe and classify MS lesions, revealing them to be distinctly demarcated focal inflammatory areas with primary demyelination, variable axonal loss and reactive gliosis in white and grey matter. Based on the presence of inflammatory cells,[20] and demyelinating changes[21] MS lesions are commonly classified as active, mixed active/inactive, and inactive.[22]

Active lesions typically develop around a central vein with a perivascular inflammatory infiltrate (T cells, B cells, monocytes/macrophages and a few plasma cells).[23,24] The surrounding lesion area is characterised by myelin loss, edema, a variable degree of axonal transection, a dense infiltration of macrophages, and some dispersed T cells.[20,24] In mixed active/inactive lesions, the lesion centre becomes less cell dense, with activated microglia and macrophages found at the lesion border.[20,24] Both active and mixed active/inactive lesions may have areas with ongoing demyelinating activity, characterised by macrophages or microglia containing myelin immunopositive for myelin protein antigens. Mixed active/inactive lesions with ongoing demyelination found at the lesion border are often called slowly expanding or smouldering lesions.[22] Inactive MS lesions are sharply demarcated, and characterised by few inflammatory cells. Almost no oligodendrocytes are present within the lesion.[22,25] In these lesions, accumulated axonal loss is often considerable,[24,25] together with astroglial activation and scar formation.[26]

MS lesions may occur in any white matter (WM) or grey matter (GM) region of the CNS. For WM lesions, predilection areas are the corpus callosum and other periventricular and juxtacortical white matter, the optic nerve, pons, cerebellum and

spinal cord.[27] Although demyelination in the cerebral cortex was described more than a century ago,[28] more careful and systematic explorations of GM lesions have not been conducted until the last 25 years.[29] Cortical lesions are commonly classified according to demyelination patterns: type I lesions involving both subcortical WM and the cortex, type II lesions being confined to the cortex, type III lesions extending from the pial surface into the cortex, and type IV lesions extending throughout the full width of the cerebral cortex.[30,31] Cortical lesions may display extensive demyelination, axonal and neuronal loss,[30] but compared to WM lesions, they appear less inflammatory, with few infiltrating lymphocytes,[30,32] less complement activation[33] and rarely BBB breakdown.[34] The most commonly occurring cortical lesion type is the type III subpial lesion. They are associated with meningeal inflammation, consisting of diffuse infiltrates of T cells, B cells and plasma cells, in severe forms resembling tertiary lymph follicles.[35] As the distribution of the cortical demyelination is consistent with an outside-in gradient, the process is suggested to be driven by a soluble cytotoxic factor produced in the meningeal inflammatory infiltrates.[36]

1.2.2 Diffuse injury in normal appearing white and grey matter

The name of MS originates from the focal demyelinating WM lesions of the CNS, and these lesions have until recently largely dominated the focus of research, diagnosis and disease monitoring. However, the pathological changes of MS expand well beyond the lesions, into so called normal appearing white and grey matter.

The diffuse injury in the normal appearing WM of MS consists of a scattered infiltration of immune cells, microglial activation, demyelination, astrocytic scarring and axonal loss.[37,38] The axonal pathology outside of lesions is thought to result from anterograde or retrograde neuroaxonal degeneration.[39] Neurodegenerative changes are also found in normal appearing GM, characterised by a loss of neurons, axons and glial cells.[36,40-42] However, the pathological mechanisms causing these changes remains elusive, although several possible pathways have been proposed. A schematic representation of some of these pathways are shown in figure 1.

First, there is strong evidence that at least part of the GM pathology occurs secondary to inflammatory damage in the WM.[29] During inflammation, activated immune cells release excessive amounts of glutamate, toxic to oligodendrocytes, neurons and axons.[43] Activated macrophages, microglia and astrocytes also produce reactive oxygen and nitrogen species that most likely cause mitochondrial dysfunction.[44-46] Sodium channels are upregulated in demyelinated axons to maintain transduction speed.[47] While compensatory in principle, this energy demanding adaptation aggravates the total energy deficiency, eventually leading to neuronal death.[48]

Second, neurodegeneration may result from pathological processes affecting GM primarily. GM demyelination and diffuse GM injury in close proximity to the meninges (i.e., the cortex) have been associated with meningeal inflammatory infiltrates, possibly through a toxic soluble factor.[36] Another hypothesis suggests that specific neuronal cell populations may be more vulnerable to changes in ion and neurotransmitter homeostasis, and are therefore more prone to neurodegeneration.[29] Finally, the inside-out theory proposes MS to be a primary degenerative disease, possibly initiated by diffuse oligodendroglial or axonal dysfunction, followed by a secondary aberrant inflammatory response.[8]

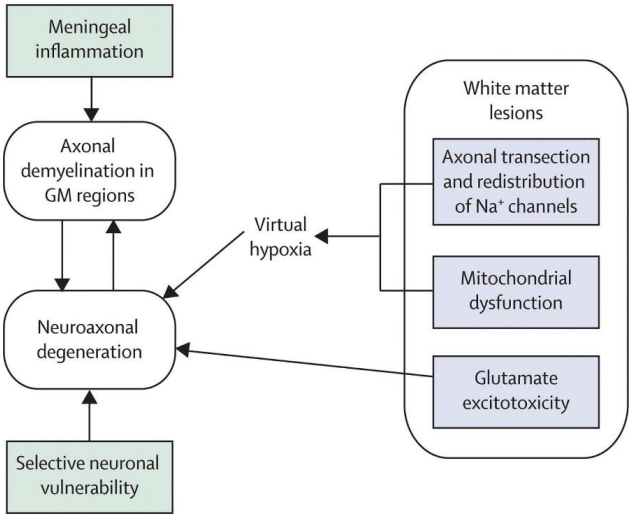


Figure 1. Schematic illustration of some of the disease processes currently believed to underlie grey matter (GM) damage in multiple sclerosis. *GM damage may occur*

secondarily to white matter (WM) pathology. Several secondary disease mechanisms (blue boxes) have been proposed, including the "virtual hypoxia" theory,[49] which postulates that axonal degeneration in the GM might result from inflammatory activity in WM lesions, combined with subsequent axonal demyelination and reorganisation of sodium channels and an inadequate mitochondrial energy supply. Alternatively, pathological processes affecting GM areas primarily (green boxes) may arise as a result of meningeal inflammation or selective neuronal vulnerability in specific predilection sites. Reprinted by permission from Elsevier: Lancet Neurol 7(9):846,[29] © Copyright 2008.

It is important to note that these mechanisms, pathways and hypotheses are not mutually exclusive. Although GM degeneration in certain brain regions seem to largely develop secondary to inflammation in the WM,[50,51] increasing evidence suggests that secondary and primary neurodegenerative processes are likely to occur simultaneously and interact with each other.[29,52] This complexity underlines the great challenge it is to clarify their separate contribution to the overall neurodegeneration.

1.3 Diagnosis and clinical course

1.3.1 Signs and symptoms

Clinical manifestations of MS consist of a wide spectrum of neurological symptoms, depending on the extent and location of lesions and diffuse CNS injury. Patients may experience unilateral blurred vision and pain with eye movements (optic neuritis), limb numbness, paresthesia and weakness (partial myelitis), or vertigo, hearing loss and double vision (brain stem syndromes).[53] In addition to sensory and motor symptoms, symptoms related to autonomic dysfunction (e.g., bladder, gastrointestinal and sexual dysfunction) and cognitive impairment (reduced attention, information processing speed, executive functioning and memory) are common, and most likely caused by pathology in complex brain and spinal cord networks.[54-56]

Upon neurologic examination, objective findings may include impaired sensation and motor weakness, ataxia, internuclear ophthalmoplegia, spasticity and hyperreflexia.

Several clinical tests and scoring systems have been developed to quantify disability in patients with MS, of which the Expanded Disability Status Scale (EDSS),[57] Timed 25-Foot Walk (T25FW) test and 9-hole peg test (9-HPT)[58] are the most commonly used. Although cognitive dysfunction affects 40-70% of patients with MS, and may severely impact their quality of life, it is still widely underdiagnosed.[55,59] Patients may experience variable deficits in several cognitive domains, which challenges the sensitivity of short and specific tests like the Paced Auditory Serial Addition Test (PASAT) and the Symbol Digit Modalities Test (SDMT) (mainly assessing processing speed). However, as larger test-batteries are often costly and time consuming, these tests are suggested as screening tools to identify patients in need of more comprehensive cognitive testing.[59]

1.3.2 Diagnosis

The diagnosis of MS is set when proof of disease activity of the CNS, disseminated in time and space, is established by clinical and paraclinical assessments. There should also be no better explanation for the clinical presentation. The paraclinical tests currently used to support the diagnosis are brain and spinal cord magnetic resonance imaging (MRI), cerebrospinal fluid (CSF) examination, and in some cases optical coherence tomography and neurophysiological testing (visual evoked potential). MRI findings (MS lesions) may determine dissemination of disease activity in both time and space, while CSF-specific oligoclonal bands may be used to demonstrate dissemination in time.[60]

1.3.3 Clinical disease phenotypes

The clinical disease course of MS varies within and between patients, and has traditionally been categorised into three main subtypes: relapsing-remitting (RR), secondary progressive (SP) and primary progressive (PP) (figure 2). Furthermore, disease activity and disease progression were included as additional descriptors in 2014.[61] Although this classification of patients is both important and practical in clinical and research settings, the phenotypes are now increasingly viewed as a continuum of clinical phases with quantitative, but not qualitative differences in immunological and neurobiological disease mechanisms.[62]

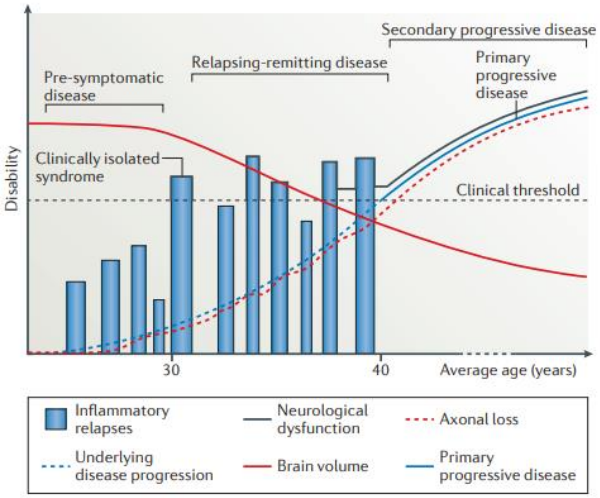


Figure 2. The heterogeneity of multiple sclerosis (MS). *Clinical disease course in patients with clinically isolated syndrome (CIS), relapsing-remitting (RR), secondary progressive (SP) and primary progressive (PP) MS according to average age and disability accumulation. Reprinted by permission from Springer Nature: Nature Reviews Immunology 15(9):546,*[48] *© Copyright 2015.*

The clinically isolated syndrome (CIS) is characterised as a monophasic clinical episode where the patient`s symptoms paired with objective findings suggest one or more inflammatory demyelinating lesions in the CNS.[60] It is considered the first clinical presentation of what later may be MS, not yet fulfilling criteria of dissemination in time.[61] The clinical episode develops over hours to days, lasting at least 24 hours, in the absence of fever or infection. The episode may subside with complete or partial recovery.[60] In CIS, active lesions in the WM are most prominent, but lesions of all types (from active to inactive) may occur in any CNS location,[62,63] as well as diffuse injury in both WM and GM.[64,65]

When monophasic clinical episodes as described above occur in patients diagnosed with MS, the disease course is characterised as relapsing-remitting. The episodes are referred to as relapses, exacerbations or attacks.[60] RRMS may be subgrouped as "active" or "not active", based on disease activity detected by clinical relapses or new

MS lesions on MRI imaging.[61] Similar to CIS, active WM lesions are the most numerous in this disease type, associated with the recurring intermittent relapses.[66] However, other lesion types, lesions in GM and diffuse GM and WM injury increase with disease duration,[62] contributing to permanent disability.[67-69]

A disease course characterised by gradually increasing neurological disability, objectively documented and independent from relapses, is defined as progressive MS. Depending on whether the disease course was progressive from disease onset, or followed a relapsing-remitting course, the terms primary progressive or secondary progressive are used, respectively.[61] By regular assessments, both disease courses may be modified by disease activity (clinical relapses or new MS lesions) and/or disease progression (progression measured by clinical evaluation).[61] In progressive disease, active lesions become rarer, while mixed active/inactive and inactive lesions are more prominent. Furthermore, cortical demyelination and diffuse GM and WM injury may be extensive.[70] Qualitatively, there are no clear differences in the pathology and immunology of SPMS and PPMS, but the proportion of active lesions and the global degree of inflammation seem to be higher in SPMS.[62]

1.4 Treatment and disease progression modifiers

The following short introduction to MS treatment will focus on disease modifying therapies (DMTs) targeting inflammatory and neurodegenerative disease processes, as well as management of important comorbidities and lifestyle factors. However, it is important to acknowledge that treatment to manage relapses, to alleviate symptoms, habilitate and rehabilitate, is essential to the quality of life and daily functioning of patients.

1.4.1 Therapeutic interventions targeting the immune response

All DMTs currently approved for MS aim to modulate, suppress or reconstitute the immune system. The principal objective is to prevent and diminish CNS inflammation, with reduced relapse rates as the main outcome of interest. Injectable DMTs (interferons, glatiramer acetate) were marketed in the mid-nineties. Reducing the

relapse rate by around 30% compared to placebo,[53] the effectiveness of these preparations is today viewed as modest.[71] From around 2005, a continuing wave of new DMTs has made more effective and personalised MS treatment attainable. Oral medications (e.g., sphingosine-1-phosphate receptor (S1PR) modulators, teriflunomide, fumarates, cladribine) have shown a modest to moderate reduction of relapse rate of 30-60% compared with placebo, while monoclonal antibody infusions (e.g., natalizumab, ocrelizumab, ofatumumab, rituximab, alemtuzumab), and mitoxantrone (now rarely used due to dose-related cardiomyopathy) are viewed as highly effective, reducing the relapse rate with 50-70% compared with placebo or active comparators.[53,71]

Anti-inflammatory in principle, the DMTs have shown clinically significant effects mainly in RRMS, and some to a smaller degree in SPMS with inflammatory disease activity.[53,71,72] For patients with PPMS, the only drug approved for treatment is ocrelizumab (anti-B cell (CD20) monoclonal antibody), associated with lower rates of clinical and MRI progression, compared with placebo.[73] In general, the effect of DMTs in progressive disease types seems to be largest for patients of younger age, shorter disease duration and more active inflammatory disease activity, while older patients without lesion activity experience little to no effect of current DM treatment.[74]

1.4.2 Therapeutic interventions targeting neurodegenerative disease mechanisms

Permanent disability in patients with MS is caused by loss of neurons and/or axons. Irreversible neuroaxonal damage during acute inflammation may leave the patient with permanent sequelae after a relapse, while the gradual, relentless disability progression independent of relapses is thought to be largely caused by neurodegenerative disease mechanisms. The latter does occur in all disease stages, but seems to dominate as a cause of increased disability in progressive disease,[74] leaving large groups of patients without effective treatment.

Against this background, neuroprotective treatment is an emerging area in MS. Treatments against specific targets in different neurodegenerative pathways (e.g., glutamate modulating agents, sodium and calcium channel inhibitors, serotonin

reuptake inhibitors) have shown promising results in animal models and early human studies, but have yet to demonstrate clinically significant reduction in disease progression.[75] Targeting only a few specific neurodegenerative mechanisms seems therefore not to provide sufficient neuroprotection, suggesting that a wider combination of targets may be necessary. To achieve this, a complete overview of the different neurodegenerative processes at play, and how they may change during the disease course, will be crucial.

1.4.3 Disease progression modifiers

Several comorbid conditions are associated with worsened prognosis in patients with MS.[53] Psychiatric disorders such as depression and anxiety are common, but often underdiagnosed, and are related to higher disability in patients.[76,77] Vascular comorbidities, including diabetes, hypertension, heart disease and peripheral vascular disease are also known to increase the risk of disability progression.[78,79] While the underlying mechanisms of these associations are unclear, and likely multifactorial in nature,[53] simultaneous treatment of MS and any comorbid conditions should be endeavoured.

Some environmental exposures and health behaviours are not only associated with increased risk of developing MS,[12,14,15] but may also impact the disease course. Patients with low levels of vitamin D have an increased risk of experiencing disease activity and progression.[80,81] Furthermore, higher body mass index (BMI) and abdominal obesity are in some studies associated with higher disability.[82,83]

There is increasing evidence that cigarette smoking affects prognosis negatively in patients with MS. Smoking is associated with disease activity, in terms of higher gadolinium-enhancing (Gd+),[84,85] T1[84] and T2[86] lesion loads in patients with CIS[85] and MS.[84,86] However, the association with lesional activity has not been consistent. In two studies using levels of cotinine, a nicotine metabolite, to define smoking status, smoking was not associated with T2 lesion volume change,[87] the cumulative number of new active lesions,[87] or the occurrence of new Gd+ and T2 lesions.[88] Furthermore, a dose-dependent relationship between cotinine levels and MRI activity was absent in

both studies.[87,88] Cotinine is considered a reliable biomarker of recent tobacco use.[89] As smoking prevalence based on patient self-reporting tends to be underestimated, cotinine levels may provide a more objective and accurate measure of tobacco use.[90] It is important to note that cotinine levels are also elevated with use of smokeless tobacco, pharmaceutical nicotine and exposure to secondhand smoke. Smokeless tobacco and nicotine have been shown not to increase the risk of developing MS.[91,92] Hence, distinguishing sources of nicotine when using cotinine levels to define smoking habits is a necessity.[93] Another marker of disease activity, higher relapse rates, are in some studies associated with smoking.[94,95] In other studies, this association has not been found.[85,87,88,96,97]

Smoking has been linked to disease progression, with several studies observing smoking patients with MS having a greater risk of converting to progressive disease types,[86,93,98-100] although the relation needs further confirmation according to recent meta-analyses.[101,102] Similarly, smoking has in some studies been associated with higher disability and disability progression,[84,86,96-98,103] while not in others.[85,87,88,104,105]

Brain atrophy, as a measure of neurodegenerative status and conceptually related to MS disease progression,[106] is associated with smoking in cross-sectional studies.[84,86,107,108] However, the longitudinal association,[85-87] as well as the association with total[84,107,108] and cortical[107] GM atrophy, is more variable.

The pathological mechanisms by which smoking may influence MS prognosis is unclear. One of the main hypotheses proposes heated cigarette smoke to induce inflammation in the lungs, triggering autoimmune reactions by sequestered self-antigens and foreign antigens in the smoke, in genetically predisposed individuals.[109,110] Furthermore, cigarette smoking may cause BBB disruption,[111] facilitating self-reactive immune cells to enter the CNS. In addition to proinflammatory mechanisms, compounds in cigarette smoke may also damage neuronal and glial cells directly, as exposure to free radicals, cyanide and oxidative stress may lead to cell death through mitochondrial dysfunction, calcium accumulation and glutamate exitotoxicity.[109]

Establishing the relation between smoking and disease activity and progression is necessary to properly advise and treat patients with MS. Studying these associations may also bring important insights to the underlying causal mechanisms, whether primarily inflammatory, neurodegenerative, or both.

1.5 Biomarkers of disease activity and progression

1.5.1 Imaging biomarkers

Lesional measures

MS lesions on MRI have served as the principal imaging marker of disease activity, and have been incorporated into current diagnostic criteria,[60] treatment goals[112,113] and outcome measures in treatment trials. By conventional MRI sequences, lesions are visualised as hyperintense or hypointense focal areas in T2- and T1-weighted images, respectively.[114] In T1-weighted images obtained after administration of gadolinium-based contrast-agents (GBCAs), lesions with ongoing inflammation and BBB disruption become bright or hyperintense, as the contrast agent leaks into the brain parenchyma.[115] Approximately 60 to 70% of WM lesions seen on histopathological examination are also identified on MRI scans with 1.5 Tesla (T) field strength, while only 5% of cortical lesions, and 15 to 40% of deep GM lesions are seen.[116]

While lesion activity has been, and still is, weighted heavily in diagnostic considerations and disease monitoring, there is a gap between the extent of MRI recognized brain lesion load and the clinical expression of the disease, popularly termed the clinico-radiological paradox.[114,117] Resolving this mismatch is important to find more sensitive and specific markers of disease activity and progression that reliably reflect the different ongoing disease mechanisms. In this search, great efforts have been made to improve how we measure both clinical disability and pathology in the CNS. First, frequently used disability rating scales like the EDSS have known limitations, including low sensitivity to change in disability, low inter-rater agreement, and a nonlinear scale.[118] To capture disability in a wider spectrum of CNS domains, the use of composite scales has been proposed.[114,117] Second, conventional lesion

imaging lacks histopathological specificity, as hyperintense lesions on T2-weighted images may contain varying degrees of demyelination, remyelination and axonal damage.[117] Hence, lesional measures with a clearer pathological substrate may provide a closer association with clinical status. Possible candidates are T1-hypointense lesions or "black holes", cortical lesions and more recently: mixed active/inactive or smouldering lesions, found in all disease types and associated with clinical disability progression,[69,119,120] However, implementation of these measures in clinical and research settings is so far limited, as they often need specialised MRI sequences, or higher field strengths for identification.[27]

Measures of diffuse tissue injury

Increased acknowledgment of extralesional MS pathology coupled with fast improvement in MRI technology over the last decades have inspired researchers to explore the relation between clinical measurements and non-lesional MRI measures. The growing range of MRI techniques may broadly assess macrostructure (e.g., brain and GM volume), microstructure (e.g., degree of neuronal damage and demyelination by diffusion tensor (DT) and magnetization transfer (MT) imaging), metabolism (e.g., axonal viability by proton magnetic resonance spectroscopy (MRS)) and neuronal function by functional MRI (fMRI).[121] Advanced imaging techniques may provide a more complete status of lesional and non-lesional tissue structure and function, as well as important insights into underlying pathological mechanisms, necessary to completely bridge the association to clinical disability. However, such comprehensive examinations require extensive resources, challenging the availability in most clinical trials, let alone in routine clinical practice.

In contrast, brain and GM atrophy measurements are obtainable from conventional MRI sequences and are thus the most widely researched macrostructural non-lesional outcome measures. Although atrophy is viewed as the end stage of neurodegeneration, reflecting irreversible neuronal and axonal loss,[122] it is found throughout the disease course, even in the earliest stages.[123,124] Furthermore, atrophy progression does not develop uniformly across tissue types, GM seems to deteriorate faster than WM,[125] and

at different rates and in different regions during the disease course.[52,125-127] Thus, studying global and regional GM measures may be more sensitive, and more clinically relevant than whole brain measures.[114]

GM atrophy may predict conversion from CIS to MS,[128] and from RRMS to SPMS.[125] It is consistently associated with disability progression in all MS disease types,[129,130] in addition to neuropsychological outcomes like cognitive impairment,[131,132] fatigue[133] and depression.[134] Due to these close relations to a wide range of clinical deficits, and with some deficits closer relations than WM lesion measures,[135] some have suggested that brain or GM atrophy measures should be incorporated into treatment goals.[113] Current treatment goals used in both research trials and clinical practice are often defined as no evidence of disease activity (NEDA), mainly emphasising the absence of inflammatory activity (occurrence of relapses and new MRI lesions), thereby largely leaving out ongoing neurodegenerative damage.[113] Incorporating MRI atrophy measures into treatment goals would provide a more balanced overview of a patient`s overall disease status, and naturally, the effect of therapeutic interventions.

Before GM atrophy can be routinely measured in individual patients, important challenges need to be addressed,[136,137] some of which are summarised in figure 3. First, technical variability related to MR image acquisition and analysis hinders the development of standardised reference values. Although brain atrophy measures are usually obtained from unenhanced 3D T1-weighted images, small differences in acquisition parameters, use of different scanners (even with identical acquisition parameters) and image analysis methods may cause systematic differences in extracted measurements.[137] Furthermore, in suggested standardised brain MRI protocols for MS, unenhanced T1-weighted images are not mandatory,[138] leaving post-contrast images often to be prioritised, especially in clinical settings where detection of inflammatory disease activity is important. Knowledge of how MRI contrast-agents may influence atrophy measurements is currently sparse,[139,140] thus limiting the use of a considerable source of real-world data.

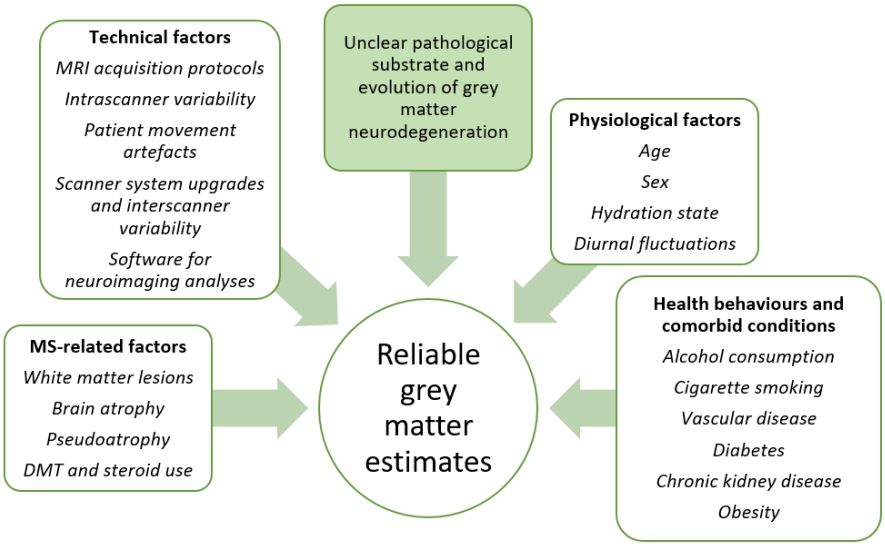


Figure 3. Challenges in the quantification and interpretation of brain grey matter atrophy in patients with MS using MRI. *Appropriate management of factors influencing grey matter volumes (transparent boxes) and underlying pathology (green box) causing variability in brain atrophy measurements is needed to ensure reliability of grey matter estimates. DMT, disease modifying therapy; MS, multiple sclerosis.*

Second, physiological variability and comorbid conditions have been shown to influence volume measurements. Brain volumes may vary with food intake,[141] body fat,[142] hydration status,[143] and time of day.[144] Moreover, established cardiovascular disease,[145] as well as risk factors of cardiovascular and cerebrovascular disease (e.g., hypertension, hyperlipidemia, smoking)[146] are associated with lower brain volumes.

Third, MS related factors may cause variability in GM measurements. Hypointense WM lesions may be segmented as GM due to the similar signal intensity in T1-weighted images, causing an overestimation of GM. Furthermore, possible effects on GM measurements caused by focal GM lesions and atrophy itself also need further investigation.[137] Another important factor still not sufficiently clarified is how DMTs may affect brain volumes, both in the short and long term. After DMT initiation, volume decrease (mainly in the WM) is seen as a result of resolved inflammation and

edema.[147,148] Beyond this initial "pseudo-atrophy" effect, most DMTs have been found to reduce atrophy rates compared to placebo in patients with RRMS,[149] although in comparative studies, highly effective treatments seem to have a more significant and long lasting effect.[150,151] The mechanisms of which DMTs with mainly anti-inflammatory effects may slow atrophy progression are largely unknown,[137] and highlight the pressing issue of untangling the possible pathways leading to neurodegeneration. If GM atrophy is to be used as an outcome measure to evaluate treatment effect, of neuroprotective treatments in particular, a clearer understanding of the extent of atrophy attributed to the different disease mechanisms will considerably strengthen the reliability of the biomarker.

1.5.2 Molecular biomarkers

Molecular biomarkers of MS disease activity classically include markers reflecting inflammatory activity, corresponding to clinical relapses or new MRI lesions in patients with RRMS. Additionally, disease activity biomarkers may also include markers reflecting ongoing neurodegeneration, measured by the rate of disability or brain atrophy progression.[152] While the latter form of disease activity becomes more apparent in progressive disease types, it is also present in relapsing disease.[62]

An ideal biomarker should qualitatively and quantitatively capture specific disease processes, a challenging task given the highly complex nature of MS pathophysiology. As the neurodegenerative aspects of MS are so closely related to disability accumulation and disease progression, the development of neurodegenerative biomarkers is often preferred, potentially serving as targets of new neuroprotective drugs, or as outcome measures when evaluating the effect of such treatments.[152] Among an increasing number of biomarkers, from the still exploratory to those used in routine clinical care, markers associated with glial activation and dysfunction, remyelination and repair, and neuroaxonal damage are often termed neurodegenerative.[152] However, given the interwoven relations between various disease mechanisms in MS, making a clear distinction between neurodegenerative and inflammatory biomarkers is challenging.

Neurofilaments

Neurofilaments are cylindrical cytoskeletal proteins, specific to neurons. They provide support to the neuronal architecture, and are thus abundantly expressed in larger myelinated axons.[153] The neurofilament consists of the subunits neurofilament light (NfL), neurofilament middle (NfM), neurofilament heavy (NfH) and α-internexin,[153] (figure 4) of which NfL is the most reliably measured[154] and most widely researched.

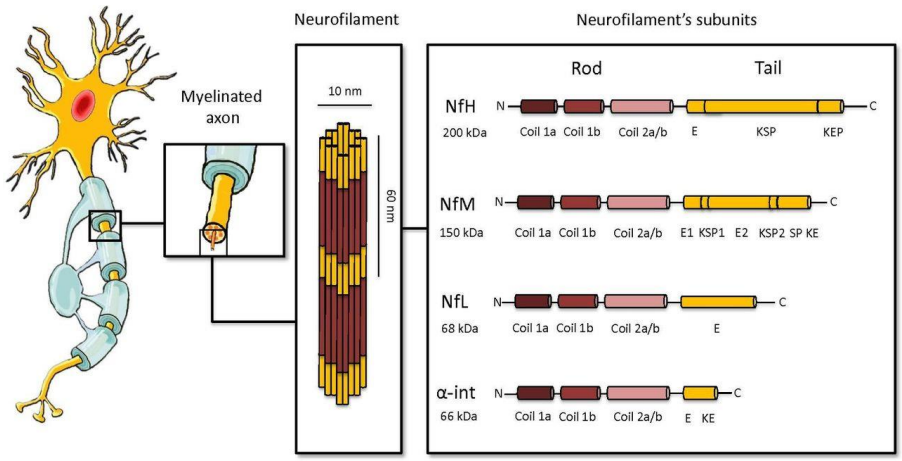


Figure 4. The structure of neurofilaments. *All of the subunits have a conserved α-helical rod domain with a variable amino-terminal and carboxy-terminal region. The length of these latter confers a different molecular weight. Neurofilament heavy chain (NfH) has the highest molecular weight and presents, in its tail, a glutamic-acid-rich segment (E segment), multiple lysine-serine-proline (KSP) repeats that are phosphorylated and a lysine-glutamic acid-proline (KEP) segment. Neurofilament middle chain (NfM) has a shorter tail with two E segments (E1 and E2), two KSP repeat segments and a serine-proline (SP) and lysine-glutamic acid (KE) segment. The tail of neurofilament light chain (NfL) is made of an E segment. Finally, α-internexin (α-int) has, in its tail, an E segment and a KE segment. Reprinted by permission from BMJ Publishing Group Ltd: J Neurol Neurosurg Psychiatry 90:871,[154] © Copyright 2019.*

Upon axonal or neuronal damage, NfL is released into the CSF and reaches the peripheral blood flow via CSF drainage into venous blood, or by diffusing through the

BBB.[155] Although NfL levels in blood are around 40-fold lower than in CSF, the correlation between CSF NfL and serum NfL (sNfL) is strong.[156] In CSF, NfL was first detected in patients with neurodegenerative diseases (including MS) in 1989 by immunoblot and enzyme-linked immunosorbent assay (ELISA) techniques.[157] However, the sensitivity of these methods is not sufficient to measure NfL levels in blood.[158] Given the invasive nature of CSF sampling, quantification of NfL in blood is necessary to achieve clinical applicability of the biomarker. Recently, widespread measurement of NfL was made possible by the development of single-molecule array (Simoa) technology,[159] capturing NfL on paramagnetic microbeads and single antibody-antigen complexes in microwells.[160] The platform has so far performed well in terms of analytical variability,[161] which together with the preanalytical stability[162,163] of sNfL considerably increases the clinical applicability of the biomarker.

At a group level, sNfL levels are consistently related to inflammatory disease activity. Higher sNfL levels have been associated with the occurrence of clinical relapses[156,164,165], the presence and number of Gd+ lesions[156,164-166] and the occurrence and number of new T2 lesions.[164-167] The association with lesion activity is especially strong for Gd+ lesions,[168] where elevated levels of sNfL may occur up to one month before,[166] and three months after a Gd+ lesion is detected on MRI.[169] These findings strengthen sNfL as a biomarker of acute neuroaxonal damage, most evidently during periods of active inflammation. This notion is further substantiated by several studies observing a significant reduction of sNfL levels after initiation of immunomodulatory drugs.[156,164,166,170] Moreover, the decrease in sNfL levels seems to be larger in response to highly effective DMTs, than less effective DMTs.[164,171]

There is currently no consensus on the association between sNfL levels and disease progression, especially on whether transition from RRMS to SPMS can be predicted.[172-174] Similarly, associations with EDSS progression are also variable, and are most consistent over relatively short time periods (a few years).[168] In these studies, patients with the highest sNfL levels (above the 80th and 90th percentile of healthy controls) had a higher risk of EDSS worsening in the following year.[156,165] In studies with a follow-up of five years or more, some found that higher sNfL levels predicted

EDSS-progression,[172,173,175] while others did not.[176,177] This variability is suggested to be influenced by the disease severity in the study cohorts (i.e., associations more likely to be found in patient cohorts with aggressive disease courses).[168] Nevertheless, it may support the use of sNfL as mainly a biomarker of inflammatory disease activity, related to acute neuroaxonal damage and short-term disability progression, rather than to neurodegenerative disease mechanisms, which one may expect to be more strongly associated with long-term prognosis.

In contrast, associations between sNfL levels and brain atrophy measures have been found in several studies.[165,172,176-179] However, studies with extensive follow-up period (ten years or more) are few, and have only considered whole brain measurements.[176,177] To better understand the relation between sNfL and atrophy progression, investigating regional GM development is necessary. This may not only clarify the value of sNfL as a neurodegenerative biomarker, but also provide further insights into underlying neurodegenerative disease mechanisms.

For both group-level comparisons and individual use, sNfL is still hampered by the lack of accepted reference values. Physiological and pathological factors may influence blood NfL levels, as illustrated in figure 5. Levels increase physiologically with age,[180] and are lower for higher BMI.[181] Furthermore, any comorbid condition causing neuroaxonal damage may influence the measured sNfL levels. Elevated sNfL levels are found in patients with other central and peripheral neurological conditions (e.g., neurodegenerative diseases,[182-184] stroke,[185] and peripheral neuropathies[186]), traumatic brain injury,[187] cardiovascular risk factors (e.g., hypertension,[188] diabetes[188] and white matter disease[189]) and can also occur due to iatrogenic causes (e.g., lumbar puncture[190]). Developing sNfL reference values ideally controlling for multiple physiological and comorbid factors have thus proved challenging. However, suggested sNfL percentiles and z-scores derived from a large reference database were recently published.[191] In that study, sNfL percentiles and z-scores were adjusted for age and BMI, and showed promising predictive value of both disease activity and disease progression.[191]

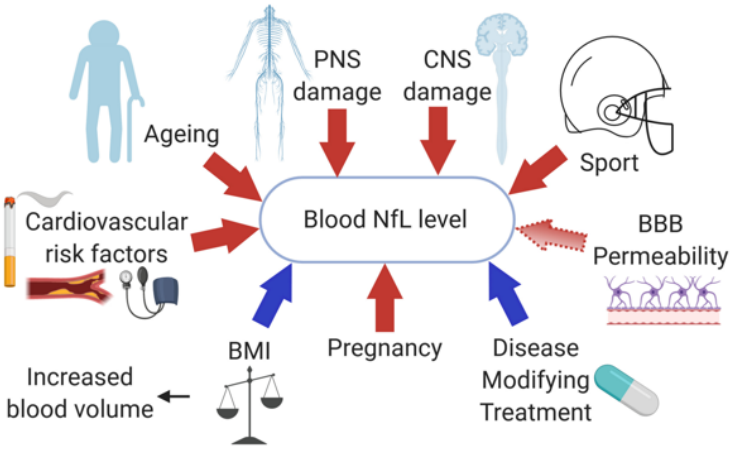


Figure 5. Factors influencing blood NfL levels. *Factors increasing (red arrows) or decreasing (blue arrows) blood levels of NfL. BBB, blood-brain barrier; BMI, body mass index; CNS, central nervous system; NfL, neurofilament light chain; PNS, peripheral nervous system. Reprinted by permission from John Wiley and Sons: Ann Clin Transl Neurol 7(12):2510,*[192] *© Copyright 2020.*

Other potential neurodegenerative biomarkers

Of biomarkers considered to be associated with neurodegenerative disease processes in MS, NfL is by far the most researched. Nevertheless, other biomarkers are considered promising, validated in terms of compatible findings in several studies, and in studies on different patient cohorts.[152]

Glial fibrillary acidic protein (GFAP) is an intermediate filament highly expressed in astrocytes. Elevated GFAP levels are associated with astrogliosis, as it is released into the CSF upon glial activation. Clinically, GFAP levels have been associated with disease progression.[193-195] Similarly, nitric oxide (NO) metabolites are found in reactive astrocytes in MS lesions,[196,197] and may be an important mechanism of axonal damage,[198] predicting disease activity[199] and disability progression.[198] Lastly, some recent studies have suggested that combining biomarkers of neuroaxonal damage (e.g., NfL) with markers of glial activation (e.g., GFAP[195]), intrathecal inflammation[200] or grey matter pathology, may increase the sensitivity in predicting disease progression.[168]

2. Study rationale and objective

2.1 Rationale

Neurodegeneration is a known negative prognostic factor in MS. Thus, it is crucial to identify its modifiable environmental risk factors, and to develop reliable and clinically useful neurodegenerative biomarkers.

Clinical implementation of candidate biomarkers is challenged by several factors. First, the pathological substrate of neurodegeneration is unclear, especially as to which degree neurodegenerative changes occur secondary to inflammatory damage in the WM, or are due to processes affecting the GM primarily. Second, variability in measurements due to technical, physiological and disease related variations, comorbid conditions and environmental factors needs clarification.

The aim of this thesis was to meet some of these challenges, by 1) untangling neurodegenerative disease mechanisms, focusing on the relationship between inflammatory WM lesions and GM atrophy, 2) assessing how technical variability related to MRI contrast agents may affect brain atrophy measurements, 3) assessing the effect of physiological and disease related variability in sNfL levels on long-term neurodegenerative and clinical disease activity, and 4) assessing the effect of smoking and cotinine levels on long-term neurodegenerative, inflammatory and clinical disease activity (figure 6).

2.2 Objectives

The objectives of this study were to:

1) Explore the spatio-temporal relationship between WM lesions and global and regional GM atrophy in patients with MS, and if the association differs in clinical phenotypes.
2) Investigate whether reliable brain atrophy measurements can be obtained from 3D T1-weighted images acquired after administration of GBCAs, using FreeSurfer.

3) Investigate how sNfL levels measured during, and outside of periods of evident inflammatory activity, associate with GM atrophy and clinical disability ten years later in patients with RRMS.
4) Assess whether smoking and serum cotinine levels in patients with RRMS relate to GM atrophy, lesion load and clinical disability after ten years.

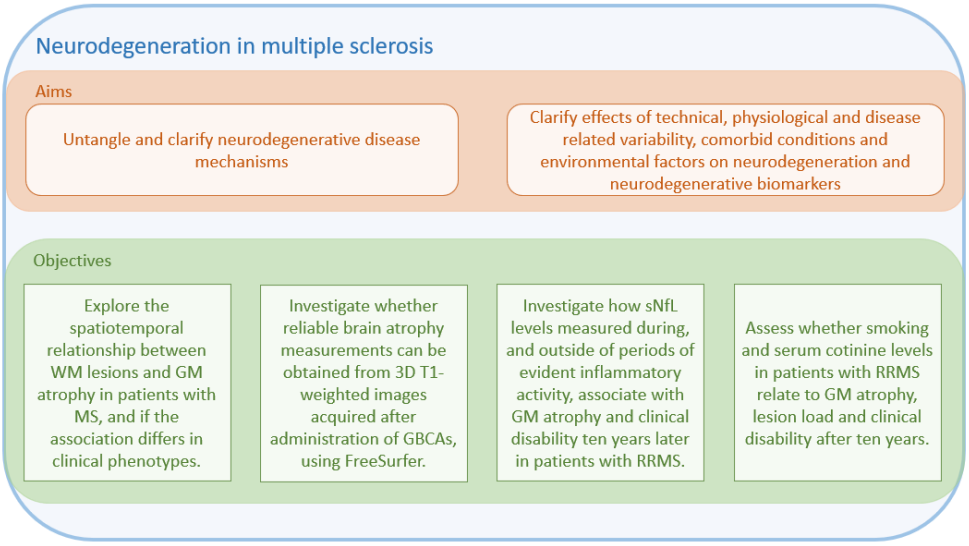


Figure 6. Illustration of the aims and objectives of the thesis. *GBCA, gadolinium-based contrast-agent; GM, grey matter; MS, multiple sclerosis; sNfL, serum neurofilament light; RRMS, relapsing-remitting multiple sclerosis; WM, white matter.*

3. Methods

3.1 Source of data: Systematic review

3.1.1 Study design and search strategy

The systematic review presented in article I was conducted and presented according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[201]

The primary outcome measures of interest were direct associations made between brain WM lesion and GM atrophy measures. A systematic search of MEDLINE (through PubMed) and Embase prior to August 17, 2020 was performed to identify relevant articles. We included studies that fulfilled all of the 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.

3.1.2 Selection, data extraction and quality assessment

Abstracts were screened by two independent raters for eligibility. According to a customised check list, extraction and quality assessment of relevant data from the included articles were conducted independently by at least two reviewers. Furthermore, the quality and risk of bias were systematically evaluated using the Quality Assessment Tool for observational cohort and cross-sectional studies (NIH, Bethesda, MA).

3.2 Source of data: The OFAMS study and the OFAMS 10-year follow-up study

3.2.1 Study design and follow-up

Research articles II, III and IV are based on data from a cohort of Norwegian patients with MS, participating in a study on ω-3 fatty acids in MS (the OFAMS study)[202] and in the OFAMS 10-year follow-up study.[104]

The OFAMS study was a multicentre, randomised, placebo-controlled trial of ω-3 fatty acids in RRMS. A detailed description of the study is previously published.[202] A total of 92 patients with RRMS[203] were included from a total of 13 neurological centres in Norway between December 2004 and July 2006. Treatment naïve patients aged 18 to 55 years with recent disease activity (≥1 relapse, or new T1Gd+ or T2 lesions on MRI within a year prior to inclusion) and an EDSS score of 5.0 or less were included. Patients with considerable somatic or psychiatric comorbidity, or whose disease activity was deemed too severe to allow delayed initiation of DMTs, were excluded.

For the first six months of the OFAMS study, patients were randomised to either ω-3 fatty acids monotherapy or placebo. From month six, both treatment groups received additional treatment with subcutaneous injections of interferon beta-1a, 44µg, thrice weekly for the remaining 18 months of the trial.

During the 24 months of the OFAMS study, six patients were lost to follow-up. One patient at baseline, three at study month two, one at month 12 and one at month 18.

The results of the OFAMS study showed no differences in disease activity (new MRI lesions, relapse rate) or disease progression (EDSS increase) between the groups receiving ω-3 fatty acids and placebo.

In 2017, *the OFAMS 10-year follow-up study* was conducted, approximately ten years after the conclusion of the OFAMS study. Of the included patients in the OFAMS study, 91 (one patient deceased) were invited to participate in the 10-year follow-up, of which 85 (93.4%) accepted. The study consisted of a clinical visit at one of the 13 study centres, as well as a brain MRI scan and blood sampling.

Between the OFAMS study and the OFAMS 10-year follow-up study, the patients had received treatment and monitoring as part of routine care. Therapeutic interventions and DMT use between the two studies were categorised based on the potency of the intervention (similar to the categorisation proposed in a recent study[191]): 1) only used platform compounds (interferon beta and glatiramer acetate preparations), 2) ever used oral therapies (teriflunomide, dimethyl fumarate, fingolimod) and 3) ever used high

efficiency monoclonal antibody therapies, chemotherapies, or haematopoietic stem cell therapy.

3.2.2 Measurements

Clinical, radiological and biochemical measurements studied in articles II, III and IV were obtained during OFAMS study visits and at the OFAMS 10-year follow-up visit, as summarised in table 1.

	The OFAMS study													**OFAMS 10-year follow-up**
Treatment	ω-3 fatty acids or placebo						ω-3 fatty acids + interferon beta-1a or placebo + interferon beta 1a							RC
Visit	BL	M1	M2	M3	M4	M5	M6	M7	M8	M9	M12	M18	M24	Y10
EDSS	●						●				●	●	●	●
T25FW	●						●				●	●	●	●
9-HPT	●						●				●	●	●	●
PASAT	●						●				●	●	●	●
SDMT														●
Serum NfL	●			●			●				●		●	
Serum Cotinine	●						●				●	●	●	
MRI	●	●	●	●	●	●	●	●	●	●	●		●	●
Self-administered questionnaire														●

9-HPT, 9-hole peg test; BL, baseline; EDSS, Expanded Disability Status Scale; M, month; MRI, magnetic resonance imaging; NfL, neurofilament light; PASAT, Paced Auditory Serial Addition Test; RC, routine care; SDMT, Symbol Digit Modalities Test; T25FW, Timed 25-Foot Walk; Y, year.

Table 1. Clinical, radiological and biochemical measurements obtained during the OFAMS study and the OFAMS 10-year follow-up study.

Laboratory tests: Serum samples were collected by venepuncture during the planned study visits, for routine analyses and cryopreservation at -80°C until post-study analyses. All serum analyses were performed blinded for patient ID and clinical data.

NfL: Serum NfL levels were measured in duplicates, from samples collected at baseline (BL), and months 3, 6, 9, 12 and 24 during the OFAMS study, using a Simoa assay and according to the manufacturer`s instruction (Quanterix, Billerica, USA).[166] Intra- and inter-assay coefficients of variation were below 10%.

For each patient, three separate mean sNfL levels were calculated: 1) "overall mean sNfL level", from all samples collected between BL and month 24, 2) "mean inflammatory sNfL level", from samples collected within two months after the appearance of a Gd+ lesion, or less than two weeks before the appearance of a Gd+ lesion, and 3) "mean non-inflammatory sNfL level", from samples collected more than two months after the appearance of a Gd+ lesion and more than two weeks before the appearance of a Gd+ lesion.

Cotinine: Serum cotinine levels in blood samples collected at BL and months 6, 12, 18 and 24 during the OFAMS study were simultaneously measured using liquid chromatography tandem mass spectrometry at Bevital AS (Bergen, Norway).[88] The lower limit of detection was 1 nmol/L, the within-day coefficient of variation was 2.0 to 6.6%, and the between-day coefficient of variation was 3.9%.

Serum cotinine levels >85 nmol/L indicate recent tobacco use[90] and are regarded to distinguish tobacco users from non-tobacco users in the general population.[204] According to serum cotinine levels, we defined smokers as patients with serum cotinine level >85 nmol/L in ≥60% of the samples, and non-smokers as patients with serum cotinine levels ≤85 nmol/L in ≥60% of the samples. To distinguish smokers from patients using smokeless tobacco, patients who reported use of snuff or other types of smokeless tobacco, but no smoking for the last ten years, were defined as non-smokers.

Self-administered questionnaire: Patients included in the OFAMS 10-year follow-up study answered a questionnaire, including habits of tobacco use and comorbid conditions.

Smoking status: We defined patients who reported to have smoked regularly within the last ten years as smokers. Non-smokers were defined as patients who did not report regular smoking, or reported use of smokeless tobacco exclusively.

Risk factors and presence of peripheral, cardio- or cerebrovascular disease: As smoking is a known risk factor for vascular conditions,[205] and because these conditions are independently associated with brain imaging changes,[146] a dichotomous variable based on the presence of patient self-reported hypertension, dyslipidemia, hypercoagulable disorders or symptomatic cardio- or cerebrovascular disease was created. This variable was included as a covariate in analyses.

MRI data acquisition

The OFAMS study: MR imaging was performed at each study site using a 1.5 T MRI scanner with the standard head coil. After intravenous injection of GBCA, the imaging protocol included a 2D sagittal fluid-attenuated inversion recovery (FLAIR) (resolution: 0.98x0.98x1mm^3, echo time (TE)/repetition time (TR)=100/6000-10000 ms, number of excitations (NEX) 2, slice thickness 4 mm), 2D axial T1-weighted images (resolution: 0.49x0.49x1mm^3, TE/TR=10-20/500-750 ms, NEX 2, slice thickness 4 mm) as well as sagittal 3D T1-weighted spoiled gradient echo (fast field echo (FFE)/fast low angle shot (FLASH)) images (resolution: 0.98x0.98x1mm^3, TE/TR=4.6/20 ms, flip angle 25°, NEX 1, slice thickness 1 mm).

The 10-year follow-up visit: Imaging was performed at the different study sites, on a 3T MRI scanner if available, alternatively using a 1.5T MRI scanner, with a standard head coil. The imaging protocol included a T2-weighted 3D sagittal FLAIR (resolution: 1x1x1mm3, TE/TR/inversion time (TI)= 386/5000/1.65-2.2 ms) and a post-contrast T1-weighted 3D sagittal magnetization-prepared rapid acquisition

gradient echo (MPRAGE) sequence (resolution: 1x1x1mm3, TE/TR/TI= 2.28/1800/900 ms, flip angle 8°).

MRI data processing

Occurrence of WM lesions during the OFAMS study: T2 and Gd+ lesion count (LC) at BL, and the appearance of new Gd+ lesions during the OFAMS study was assessed by two experienced neuroradiologists, blinded for patient ID and clinical data.

Lesion segmentation and lesion filling: The following MRI processing was performed on images obtained during the OFAMS 10-year follow-up study. Lesion segmentation was done on FLAIR images using Lesion Segmentation Tool (LST) (version 2.0.15; http://applied-statistics.de/lst.html).[206] The lesion probability map in FLAIR space was brought to T1-weighted space by FMRIB`s Linear Image Registration Tool (FLIRT) linear registration of the FLAIR image to the T1 image, using 7 degrees of freedom, correlation ratio as the cost function, and trilinear interpolation. Afterwards, a threshold of 0.1 was used to binarise the lesion probability map. To optimise the lesion filling, gadolinium-enhancing regions (both lesions and other regions) were first removed, by applying an upper intensity threshold at the 98th percentile. Next, the FMRIB Software Library (FSL) (version 5.0.10; http://www.fmrib.ox.ac.uk/fsl) was used to fill in lesional voxels in the T1-weighted images using the lesion_filling tool,[207] and these filled lesions were pasted into the original post-contrast 3D T1-weighted images.

Morphological reconstruction: The following MRI processing was performed on images obtained during the OFAMS 10-year follow-up study, and on a subset of images obtained at month 24 during the OFAMS study. Cortical reconstruction and parcellation for cortical volume and thickness measurement and subcortical segmentation were performed with FreeSurfer, (version 7.1.1; http://surfer.nmr.mgh.harvard.edu/), a freely available software package for academic use, available through online download. The technical details of FreeSurfer procedures have been previously described.[208,209]

Quality control was performed by visual inspection, and any segmentation errors were recorded for each patient. In cases where only specific anatomical regions were incorrectly segmented, we chose to not apply any corrections for these errors in our analyses.

The Desikan-Killiany atlas[210] was used to extract cortical thickness measures. Furthermore, total cerebral GM and WM volume, total deep GM and thalamus volume (left and right hemisphere) were obtained.

3.2.3 Ethical approvals and patient consent

The OFAMS study (REC-WEST-ID 9481/No.:2014/1120 & 005.04) and the OFAMS 10-year follow-up study (REC-WEST-ID 17299 / No.:2016/1906) were approved by the Regional Committee for Medical and Health Research Ethics in Western Norway Regional Health Authority. All participants gave their written informed consent.

3.3 Statistical analyses

Article I

To assess the association between brain WM lesions and both global and regional GM atrophy, in the different disease phenotypes, qualitative and descriptive analyses were performed.

Article II

Paired t-test were used to assess differences in structural measurements between pre- and post-contrast measurements. Before the analyses, the data distribution was visually and statistically evaluated using the Kolmogorov-Smirnov test for normality. Bland-Altman plots were created to identify fixed or proportional bias.[211] The intra-class correlation coefficient (ICC) was determined to assess the agreement between volume and thickness measurements obtained before and after GBCA administration.

Analyses were performed using the Statistical Product and Service Solutions (SPSS) for macOS (version 25; SPSS, Chicago, Illinois).

Articles III and IV

The outcome measures of interest were MRI and clinical measurements obtained at the 10-year follow-up visit, and the change in clinical measurements from month 24 of the OFAMS study to the 10-year follow-up study.

Analyses were performed using SPSS for macOS (version 25; SPSS, Chicago, Illinois) and R software (version 4.0.5).

Article III: The relationship between mean overall sNfL and the primary outcome measures was assessed by a linear multilevel model. To correct for scanner variability, study site was entered as a random effect. Age, sex, DMT use, estimated intracranial volume (eTIV), fraction of MRI scans with new Gd+ lesions (fGd+), BL T2 and Gd+LC were included as covariates.

The association between mean inflammatory and non-inflammatory sNfL and the primary outcome measures was investigated by linear regression models, as entering study site as a random effect did not improve the model. In the models including mean inflammatory sNfL as the predictor, fGd+, age, sex, DMT use, eTIV, BL T2 and Gd+LC were entered as covariates. A modified version of this model was also used in two exploratory analyses. In the first of these analyses, MRI atrophy measures obtained at month 24 (available in a subset of patients) were included as a covariate. In the second model, mean cortical thickness in the precentral gyrus was investigated as the dependent variable. In the models including mean non-inflammatory sNfL as the predictor, age, sex, DMT use, eTIV, BL T2 and Gd+LC were included. All independent variables were first entered as covariates and removed by backward elimination if not significant to the model. Assumptions for linear regression were checked for each final model, log-linear transformation was performed if assumptions were not satisfied.

To assess the dichotomous outcome measure EDSS≥4, logistic regression models were used. Finally, to control the false discovery rate for multiple hypothesis testing, the Benjamini-Hochberg method[212] was used.

Article IV: Analyses were first performed using smoking status defined by serum cotinine levels as the predictor, then repeated using smoking status defined by patient self-reporting.

The difference in outcome measures between smokers and non-smokers was assessed by a two-sample t-test for normally distributed variables, otherwise, Mann-Whitney tests were used. Kolmogorov-Smirnov tests and visual inspection of the histograms were used to assess normality of the data distribution.

The association between smoking status and mean cotinine levels and the outcome measures was investigated by a linear multilevel model. The MRI scanner used was entered as a random effect, to correct for scanner variability. Furthermore, age, sex, presence of vascular disease, eTIV, BL EDSS, and time from diagnosis were included as covariates. For each final model, assumptions for linear regression were checked. Log-linear transformation was performed if the assumptions were not satisfied.

4. Results

4.1 Article I

A total of 90 studies were included in the systematic review, based on the screening of 2 260 unique citations. Of these, 64 studies reported cross-sectional analyses, 18 reported longitudinal analyses and 8 reported both cross-sectional and longitudinal analyses.

In the majority of studies, higher WM lesion load was associated with more global, cortical and deep GM atrophy. The association was most consistently found in cross-sectional studies, and in patients with RRMS. While the relation with WM lesions in this disease phenotype was significant in most studies considering global, cortical and deep GM, the association was particularly consistent with deep GM and thalamus volumes, which was also seen in CIS. For both SPMS and PPMS, the associations were more variable than in the relapsing disease types. In the majority of studies, higher WM lesion load was related to lower global GM volume, while the associations with cortical and deep GM atrophy were less consistent.

The frequent relation between WM lesions and all GM regions in RRMS and deep GM in CIS, suggests that early GM neurodegeneration is mainly secondary to inflammatory damage in the WM. While still present in progressive disease types, the associations were more variable, indicating that the predominating neurodegenerative disease mechanism may have shifted to a primary process.

4.2 Article II

In the 10-year follow-up study, a total of 23 patients had pre- and post-contrast T1 weighted images obtained with the identical acquisition protocol. One of these patients was excluded from further analyses, due to an image artifact causing large segmentation errors.

Minor to moderate segmentation errors were found in all images, but more frequently and severely in post-contrast images. We identified three common errors, two of which occurred during the construction of the pial surface (representing the border between cortical GM and CSF) or the border delineating deep GM structures. Both errors resulted in an overestimation of the cortical thickness and deep GM volumes. The third common error occurred during the construction of the white surface (representing the border between WM and GM). Here, the white surface failed to follow the intensity gradient correctly, resulting in a suboptimal segmentation, most frequently seen in the temporal poles.

We found good to excellent consistency between measurements obtained before and after GBCA administration, with all ICC values above 0.92. The results of the t-tests showed that GM volumes and cortical thickness measurements were systematically higher in post-contrast images, while the total WM volume decreased. While there was no proportional bias, these systematic differences were confirmed in Bland-Altman plots.

Due to the high consistency of measurements between pre- and post-contrast images, we concluded that reliable GM volume and cortical thickness measurements may be obtained from post-contrast 3D T1-weighted images, using FreeSurfer. However, the systematic overestimation of the GM means that measurements from pre- and post-contrast images should not be compared directly.

4.3 Article III

We included 78 of the 85 patients participating in the 10-year follow-up study, in which sNfL measurements were available.

The overall mean sNfL level did not predict any MRI or clinical measurements after ten years. When assessing mean inflammatory sNfL level, we found significant associations between higher levels and lower total GM and deep GM volume, lower mean cortical thickness, and higher log T2 LC after ten years. Furthermore, higher inflammatory sNfL levels predicted a higher score (higher disability) on the log

dominant hand 9-HPT. The fraction of MRI scans with a new Gd+ lesion, or the mean non-inflammatory sNfL level did not predict long-term atrophy or disability progression.

Our findings suggest that sNfL levels measured during periods of active inflammation in patients with RRMS may be a way to quantify the extent of ongoing axonal damage, predicting long-term atrophy and disability progression. The association between GM atrophy and inflammatory sNfL, but not overall sNfL or non-inflammatory sNfL level, may also imply that the subsequent GM atrophy in patients with RRMS is largely caused by neuroaxonal degradation secondary to inflammatory damage.

4.4 Article IV

All 85 patients who participated in the 10-year follow-up study was included. One patient did not have serum samples available for cotinine analyses, and another patient did not complete the questionnaire. Thus, 84 patients were categorised as non-smoker or smoker by each definition. Defined by cotinine levels, 37 patients were non-smokers, and 47 smokers. By patient self-reporting, there were 48 smokers and 36 non-smokers.

When comparing MRI and clinical measurements in non-smokers and smokers, patients with MS who were defined as smokers (by serum cotinine levels) had lower total WM and deep GM volumes, and higher T2 lesion volumes after ten years. Smokers also had a higher score (more disability) on the T25FW test, and a larger decrease in PASAT scores from month 24 to the 10-year follow-up visit. The results were similar when defining smoking by patient self-reporting.

In the linear multilevel models, smoking (defined by serum cotinine levels) was associated with lower WM volume and higher log T2 lesion volume after ten years, but not with any of the clinical measurements. When defining smoking by patient self-reporting, smoking was additionally associated with lower deep GM volume, a higher score on the log T25FW test, and a larger decrease in PASAT scores. Finally, mean

cotinine levels in smokers (defined by serum cotinine level) were not associated with any of the outcome measures.

Based on the findings, we concluded that smoking may have a negative long-term influence on atrophy and disability progression in patients with RRMS. Thus, MS patients who smoke should be offered advice and help in smoking cessation as soon as possible after diagnosis.

5. Discussion

5.1 The contribution of the findings

Relationship between white matter lesions and grey matter atrophy in MS

The findings of this thesis add to the evidence of brain GM atrophy being related to inflammatory lesions in the WM. In article I, the relation was consistent for all GM regions in RRMS, and for the thalamus and deep GM in CIS. Although still present in most studies on progressive disease types, the relationship between WM lesions and GM atrophy was more varied. The results suggest that GM atrophy in early (relapsing) MS largely develops secondary to inflammatory damage in the WM, while in progressive disease types, neurodegeneration may be dominated by primary mechanisms affecting the GM directly. Different primary and secondary mechanisms of neurodegeneration in MS have previously been discussed in numerous review articles, assessing evidence from both histopathologic and imaging studies.[6,8,29,62,213-219] However, a systematic review of the existing knowledge of the in vivo relationship between WM lesions and GM atrophy in MS, and especially how it may differ in disease phenotypes, has previously not been done. Thus, our results strengthen and further the reports of previously conducted research, by the objective and comprehensive gathering, interpretation and presentation of the existing evidence.

A strong relation between inflammatory WM lesions and GM atrophy in RRMS is further supported by the findings in article III. Here, we found significant associations between higher sNfL levels during periods of active inflammation (reflective of the extent of acute axonal damage)[169] and more GM atrophy after ten years. For sNfL levels collected during periods of remission, such associations were not found. These findings suggest that the associated GM atrophy develops secondary to inflammatory axonal damage, most likely through neuroaxonal degeneration through connected WM tracts.[48] In our study, inflammatory sNfL level was most strongly associated with lower cortical thickness and deep GM volume, regions highly interconnected through WM tracts,[220] therefore vulnerable to secondary degeneration.[221,222] This is partly in line

with previous studies,[172,178,223] two of which found an association between higher sNfL levels at baseline[172] or during the first two years of the follow-up,[223] and deep GM volume loss over five[172] and ten years.[223] In a third study, the patient group with the most prominent spinothalamic volume loss over six years also had the highest mean sNfL levels.[178] These previous studies did not discern between inflammatory and non-inflammatory sNfL levels. Our work may therefore provide further insight to the possible underlying neurodegenerative disease processes, by linking lower GM measures to the extent of axonal damage during acute inflammation.

Deep GM atrophy has previously been related to WM lesions,[123,221] and shown to progress at a faster rate than other GM regions, especially in early disease stages.[52,123] In article IV, we found that smoking patients with MS had lower deep GM and total WM volumes, as well as higher T2 lesion volumes after ten years. These findings may be in support of GM neurodegeneration occurring mainly secondary to inflammatory WM damage in RRMS, and that the detrimental effect of smoking in MS is largely mediated by increased inflammatory activity. A relation between smoking and increased lesion load has previously been shown in patients with MS,[84-86] although not consistently, as two studies using cotinine levels to define smoking status (including one study investigating patients from the same cohort as in our study) did not find a significant association.[87,88] The longitudinal relationship between smoking and brain atrophy has so far mainly been studied for whole brain atrophy, with inconsistent results.[85-87] Of the few studies assessing GM atrophy, one study found that current smoking was associated with lower grey matter fraction,[108] while the others did not.[84,107] Overall, our study suggests that smoking increases neurodegeneration in patients with MS, possibly through inflammatory disease mechanisms. It is however important to note that more studies are needed to conclude on causal mechanisms.

Effect of gadolinium-based contrast-agents on GM atrophy measurements

This thesis has contributed with new findings of reliable GM volume and cortical thickness measures obtained from post-contrast 3D T1-weighted images using FreeSurfer. These findings, reported in article II, may allow use of historic and

prospective real-world data, collected as part of routine care or research trials, as post-contrast images are often prioritised to detect inflammatory disease activity.[138] Only a few studies have previously investigated the effect of GBCAs on automated brain tissue measurement in MS, using software packages other than FreeSurfer.[139,224,225] These papers also reported good consistency between measures obtained before and after contrast administration, suggesting that post-contrast atrophy measurements may be reliable across segmentation techniques.

Another novel finding in article II was the systematic increase in GM measurements in post-contrast images, meaning that pre- and post-contrast images should not be compared directly. The definite cause of these differences could not be determined, although the higher intensity in some extraparenchymal structures in post-contrast images seemed to cause focal challenges in correctly separating different tissue types. Nevertheless, it is uncertain whether the increase in almost all GM regions and overall decrease in WM volume can be fully explained by these focal segmentation errors.

sNfL as a biomarker of long-term neurodegeneration and disability progression

By relating higher inflammatory sNfL levels to more GM atrophy and clinical disability ten years later in article III, this thesis supports sNfL as a biomarker of long-term neurodegeneration. From previous research, the role of sNfL as a marker of acute axonal damage, predicting short-term disease activity, treatment response and disability progression, has become increasingly established.[168] However, its ability to predict long-term disability and brain atrophy progression has so far been less clear.[168,226,227] The association between higher inflammatory sNfL levels and lower deep GM volume after ten years in patients with RRMS, is in line with previous studies.[172,178,223] Additionally, we found an association between higher inflammatory sNfL levels and lower mean cortical thickness, which, to our knowledge, has previously not been found. Of the clinical outcomes, higher inflammatory sNfL levels were in our study related to higher disability measured by the 9-HPT ten years later. Similar findings were reported in a previous study, where the group with higher sNfL levels developed motor disability faster, with the most evident increase in the T25FW

test and the 9-HPT.[178] In article III, associations with GM atrophy and clinical disability were specifically tied to sNfL levels measured during periods of ongoing inflammation, reflective of acute axonal damage. Contrarily, associations were not found for sNfL levels collected during periods of remission, possibly more reflective of ongoing neurodegenerative processes. Although this is the first study to investigate sNfL levels collected during periods of acute inflammation and remission separately, missing associations with non-inflammatory sNfL are in line with a recent study on natalizumab-treated patients, also finding sNfL levels not to be related to disease progression.[228] In summary, our study suggests that sNfL levels measured during (or in a limited time period before and after) relapses may indicate the risk of permanent or future disability progression, by quantifying the extent of ongoing axonal damage. Furthermore, the findings of this thesis highlight important factors to consider when biomarkers not specific to one pathologic disease mechanism are used in clinical decision making: the dynamic nature of MS pathophysiology calls for careful timing of measurements dependent on the clinical problem in question (e.g., presence of disease activity, therapeutic effect or long-term prognosis), in order to correctly interpret the results.

Effect of smoking and cotinine levels on long-term neurodegeneration and disability progression

This thesis adds to the notion of smoking being a risk factor of increased neurodegeneration in MS, possibly through heightened inflammatory disease activity.[109]

The associations between smoking and disability accrual were in article IV overall modest, only found for the T25FW test and PASAT. In previous longitudinal studies, smoking predicted worsened cognitive function measured by PASAT,[229] while associations with EDSS progression have been variable.[86,87,103,105] Nevertheless, our findings indicate an overall unfavourable effect of smoking on long-term prognosis, and supports that smoking patients with MS should be offered advice and support in smoking cessation, as part of routine care.

All analyses in article IV were conducted separately for two definitions of smoking, one based on serum cotinine levels, and one based on patient self-reporting. This has to our knowledge not been done previously. The overall comparable results suggest that serum cotinine levels provide a reliable and objective alternative to patient self-reporting, to study the possible effects of smoking.

5.2 Methodological considerations and limitations

5.2.1 Level of evidence for observational studies

The evidence level of a study is traditionally ranked according to the internal validity of the research design used.[230] Internal validity describes the study`s ability to measure what it intended to measure, namely, the correctness of the results.[231] Whereas the external validity of a study refers to the extent to which you can generalise the findings to other populations.[232] An example of the ranking is shown in figure 6, where results obtained from randomised controlled trials (RCTs) and systematic reviews and meta-analyses of RCTs are viewed as the highest quality of evidence, with the lowest risk of bias. Conceptually, bias is the lack of internal validity or any systematic error causing an incorrect estimate of the relation between an exposure and an effect in the target population. Internal validity in clinical research is paramount, as clinical decision making based on invalid research results would be both worthless and potentially dangerous.[231] Furthermore, external validity cannot exist without internal validity.[232]

The stringent methodology and random allocation of participants in RCTs are thought to eliminate systematic bias, allowing for causal inference.[233] However, RCTs are often costly and time consuming, which may result in small studies with insufficient follow-up time.[234] Furthermore, the deductive methodology carries a certain narrowness of scope, meaning limited external validity.[235] A rigid hierarchy ranking evidence from RCTs as “gold standard” has thus been challenged, with some studies suggesting that the appropriate research design in large depends on the type of research question asked.[230,234,236]

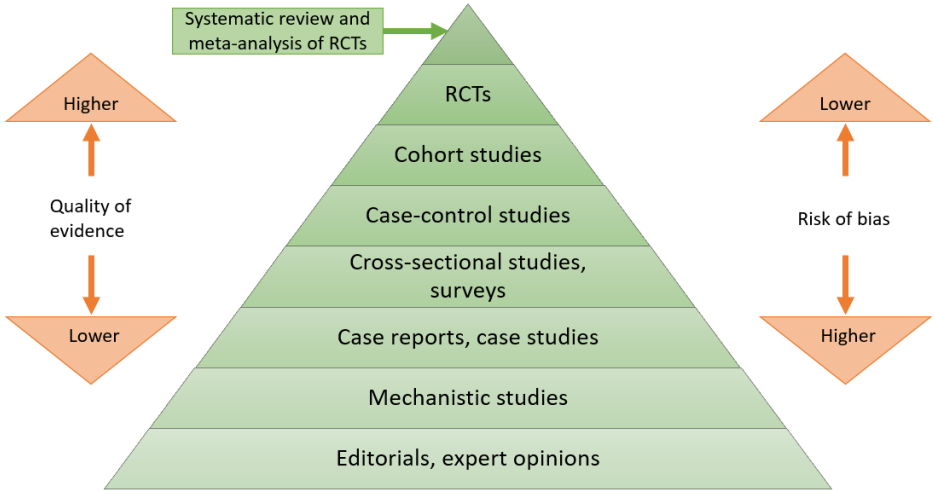


Figure 6. Hierarchical ranking of the quality of evidence according to study design. *Each category is considered methodologically superior to those below it. RCT, randomised controlled trial. Adapted and reprinted by permission from Oxford University Press: Am J Clin Nutr 105(1):249S285S,*[237]*© Copyright 2016.*

The results of this thesis are obtained from observational studies (retrospective cross-sectional and longitudinal cohort studies), and a systematic review of (mostly) observational studies, which fall at intermediate quality levels (figure 6).[230] To answer questions about prognosis, longitudinal cohort studies are well suited,[230,234] especially when blinded allocation according to exposure/non-exposure is difficult, or ethically impossible. These considerations are relevant to this thesis, where conducting an RCT would mean allocating patients to smoke or not smoke in article IV. In article III, patients would be allocated to either highly effective treatment or placebo, to investigate sNfL levels during periods of disease activity and remission. Nonetheless, all observational studies have built-in bias, which may cause spurious or indirect associations, as opposed to causal.[231] To critically assess the nature of detected associations is complicated, and entails quantifying errors (as nearly every study will have them) up against a non-existent cut-off indicating when a study should be considered invalid.[238] As part of this evaluation, it is common to assess the internal

validity of observational studies according to three categories of bias: selection, information and confounding bias.[231]

5.2.2 Selection bias and sample size

Selection bias is the error introduced when the study population does not represent the target population,[232] or when the exposed and unexposed groups differ systematically in aspects aside from the exposure.[231]

The patients studied in articles II to IV were originally recruited to an RCT on ω-3 fatty acids, adhering to strict inclusion and exclusion criteria.[202] These observational studies are therefore subject to some of the same types of selection bias commonly seen in RCTs. Specifically, participants in RCTs tend to be healthier than those who do not take part.[231] In the OFAMS study, patients with severe comorbidities and/or highly active disease were excluded, which may be reflected in the relatively low disability progression (mean EDSS progression of 0.9 points[104]) during the 10-year follow-up. Furthermore, the extensive follow-up period may have led to selection bias by differential loss to follow-up,[233] if for example smoking status, or disability level affected the participants decision to attend the 10-year follow-up visit. Considering the high attendance (93.4%), it seems however less likely that this would have affected the results significantly.

As the number of patients included in the OFAMS study was based on pre-determined power calculations and effect assumptions, the relatively small sample size may not be optimal for later observational studies on the same cohort. Increased random variation and decreased precision may occur in small sample sizes, therefore prone to type II errors (false negative findings).[239] This may have affected some analyses in this thesis, especially those concerning clinical outcome measures known to be less sensitive to change, and those on small subgroups of patients.

Lastly, many of the included studies in the systematic review (article I) did not have the relation between WM lesions and GM atrophy as their main topic of interest, possibly only including these results when they found significant associations, increasing the risk of publication bias. To minimise this effect, we selected only papers

that mentioned in their abstracts either that the association was assessed or the result of that assessment. Nonetheless, such bias could still be present.

5.2.3 Information bias

Information bias occurs during data collection, from incorrect or inaccurate measurements of key study variables, or from information obtained differently in exposed and unexposed participants.[231,232]

In retrospective cross-sectional and longitudinal cohort studies, one of the challenges for the investigator is the limited control over data collection.[240] In articles II to IV, the MRI data from the 10-year follow-up visit were obtained from up to 15 different scanners at 13 study sites. Variability in scanners and acquisition parameters is a factor known to affect brain measurements,[136,137] and although key information on MRI hardware and protocols was provided, certain details (e.g., head coil type, dose and type of GBCA used) were difficult to retrospectively retrieve for all patients. Nevertheless, as we have no reason to believe the patients were systematically allocated to certain scanners based on specific exposures, and the image analyses were done with the same methodology for all scans, it is less likely that these factors introduced considerable information bias in our studies.

In article I, included papers were published over a span of twenty years. The major improvements in MRI technology in that same period, entail a vast methodological interstudy variability, including MRI scanners and acquisitions, image (pre)processing tools and analysis software. This variability does not necessarily hurt the internal validity of the included studies, but the internal validity of the systematic review itself. To avoid this, interpretations and analyses were kept qualitative, instead of conducting meta-analyses using point and interval estimates from methodologically incomparable studies.

Lastly, by classifying sNfL levels as inflammatory or non-inflammatory in article III, and patients as smokers or nonsmokers in article IV, the occurrence of misclassification bias[232] is possible. However, all analyses were in both articles conducted twice: in article III using the mean sNfL level calculated from at least two and three

measurements; and in article IV using two separate definitions of smoking (by serum cotinine levels and by patient self-reporting). The analyses yielded highly comparable results, suggesting that misclassification bias did not affect the main conclusions of the studies.

5.2.4 Confounding bias

If present, a confounder acts as a third factor, when relating an exposure to an effect. A confounding variable is related to the exposure, it affects the outcome, but it is not an intermediate step in the causal pathway between the exposure and the effect.[231,232] While selection and information bias may cause spurious associations, confounding may cause indirect associations, which are real, but not causal.[231]

The MRI brain measurements investigated in articles II to IV were mainly extracted cross-sectionally from images obtained at the 10-year follow-up visit. This obviously limits the ability of this thesis to conclude on how GBCAs, sNfL levels and smoking affect longitudinal brain atrophy progression in patients with MS. The considerable difference in MRI scanners and acquisitions between the OFAMS study and the 10-year follow-up did not allow longitudinal analyses. Furthermore, the varying quality of the MRI images obtained during the OFAMS study led to a successful image analysis in only a small subset of patients. Thus, adjusting for baseline MRI brain measurements, or the change in those measurements over the 24 months of the OFAMS study was not possible for the main analyses in article III and IV. Although baseline lesion counts (visually determined) were included as covariates in some analyses, we cannot fully exclude that early MRI measurements confounded the results. If so, the differences in brain atrophy and lesion measures seen after ten years may have already been present at baseline, or occurred during the first two years of the follow-up.

In addition to the early MRI measurements, our findings may have been affected by other unmeasured, or partly measured confounding. For example, both therapeutic interventions and comorbid conditions are known to affect prognosis,[53,71,145] but the effect of these factors may vary considerably both between and within patients, according to the current type of therapeutic intervention or comorbid condition. In

article I, the reviewed papers did not consistently account for the effect of either therapeutic interventions, physiological variability or comorbid conditions. Together with the relatively low number of longitudinal studies included, this limits our ability to conclude on the temporal relations between inflammatory and neurodegenerative disease activity, and factors possibly modulating these relations. In article III and IV, patients had used a variety of therapies at different times and duration during the follow-up, making it challenging to statistically consider all aspects of the individual DMT use. Nevertheless, the associations found in article III between inflammatory sNfL levels and atrophy in certain GM regions were still present when adjusting for DMT use during the follow-up. This is in line with the findings in article I, where inflammatory WM lesions were consistently associated with GM atrophy in patients with RRMS, and especially with atrophy in deep GM. Overall, the results in article III and IV are consistent with previous key studies, finding a predictive value of sNfL,[165,172,177] and negative effects of smoking[86] on long-term neurodegeneration in patients with MS. Although consistent findings with previous observational studies, or within the 90 studies included in the systematic review do not exclude the possibility that unmeasured confounding may fully explain our results, it makes it less likely.

6. Future perspectives

The aim of this thesis was to investigate neurodegeneration in MS, including biomarkers reflecting these processes, and environmental factors possibly influencing them.

Exploring the clinical applicability of candidate biomarkers, this thesis has contributed with novel findings, of reliable brain atrophy measures obtained from post-contrast images using FreeSurfer. Further investigations into the effects of GBCAs on MRI atrophy measurements should also consider other brain segmentation software, and potential systematic effects dependent on the type and dose of GBCA used, and the delay time after injection.

We have also contributed with further knowledge on the relationship between inflammatory WM damage and GM atrophy by reviewing the existing literature, and by investigating how sNfL levels reflective of acute axonal damage associate with future GM loss. Similarly, we have shown that smoking in patients with MS is likely to increase long-term neurodegeneration. To further clarify the details of these (spatio)temporal relationships, and to approach the question of causality, more studies are needed. Preferably longitudinal, prospective imaging studies, with sufficient sample size and follow-up time, the appropriate image acquisitions included, and as little technical inter- and intrastudy variability as possible. In models predictive of neurodegeneration, known risk factors and modulators of neurodegeneration should be adjusted for, for example current and previous therapeutic interventions, and baseline and on-study lesion activity. Further, each neurodegenerative process needs to be assessed separately for each disease phenotype, in order to untangle primary and secondary processes, and to determine the predominating neurodegenerative disease process. Lastly, the results of these in-vivo studies should be interpreted or combined in context with knowledge obtained through molecular or histopathologic studies.

References

1. Walton C, King R, Rechtman L, et al. Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. *Mult Scler.* 2020;26(14):1816-1821.
2. Uitdehaag B, Kobelt G, Berg J, Capsa D, Dalen J. New insights into the burden and costs of multiple sclerosis in Europe: Results for the Netherlands. *Mult Scler.* 2017;23(2_suppl):117-129.
3. Ziemssen T, Derfuss T, de Stefano N, et al. Optimizing treatment success in multiple sclerosis. *J Neurol.* 2016;263(6):1053-1065.
4. Wucherpfennig KW, Strominger JL. Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell.* 1995;80(5):695-705.
5. Sospedra M, Martin R. IMMUNOLOGY OF MULTIPLE SCLEROSIS. *Annu Rev Immunol.* 2004;23(1):683-747.
6. Hemmer B, Kerschensteiner M, Korn T. Role of the innate and adaptive immune responses in the course of multiple sclerosis. *Lancet Neurol.* 2015;14(4):406-419.
7. Barnett MH, Prineas JW. Relapsing and remitting multiple sclerosis: pathology of the newly forming lesion. *Ann Neurol.* 2004;55(4):458-468.
8. 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.
9. Patsopoulos Nikolaos A, Baranzini Sergio E, et al. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science.* 2019;365(6460):eaav7188.
10. Sawcer S, Hellenthal G, Pirinen M, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature.* 2011;476(7359):214-219.
11. Bjornevik K, Cortese M, Healy Brian C, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science.* 2022;375(6578):296-301.
12. Munger KL, Levin LI, Hollis BW, Howard NS, Ascherio A. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. *JAMA.* 2006;296(23):2832-2838.
13. Munger KL, Bentzen J, Laursen B, et al. Childhood body mass index and multiple sclerosis risk: a long-term cohort study. *Mult Scler.* 2013;19(10):1323-1329.
14. Munger KL, Chitnis T, Ascherio A. Body size and risk of MS in two cohorts of US women. *Neurology.* 2009;73(19):1543-1550.
15. Hernán MA, Oleky MJ, Ascherio A. Cigarette Smoking and Incidence of Multiple Sclerosis. *Am J Epidemiol.* 2001;154(1):69-74.
16. Ghadirian P, Dadgostar B, Azani R, Maisonneuve P. A case-control study of the association between socio-demographic, lifestyle and medical history factors and multiple sclerosis. *Can J Public Health.* 2001;92(4):281-285.

17. Zhang P, Wang R, Li Z, et al. The risk of smoking on multiple sclerosis: a meta-analysis based on 20,626 cases from case-control and cohort studies. *PeerJ.* 2016;4:e1797.
18. Charcot J-M. *Histologie de la sclerose en plaques.* Paris: [s.n.]; 1868.
19. Zalc B. One hundred and fifty years ago Charcot reported multiple sclerosis as a new neurological disease. *Brain.* 2018;141(12):3482-3488.
20. Bö L, Mörk S, Kong PA, Nyland H, Pardo CA, Trapp BD. Detection of MHC class II-antigens on macrophages and microglia, but not on astrocytes and endothelia in active multiple sclerosis lesions. *J Neuroimmunol.* 1994;51(2):135-146.
21. Lucchinetti C, Brück W, Parisi J, Scheithauer B, Rodriguez M, Lassmann H. Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Ann Neurol.* 2000;47(6):707-717.
22. Kuhlmann T, Ludwin S, Prat A, Antel J, Brück W, Lassmann H. An updated histological classification system for multiple sclerosis lesions. *Acta Neuropathol.* 2017;133(1):13-24.
23. Machado-Santos J, Saji E, Tröscher AR, et al. The compartmentalized inflammatory response in the multiple sclerosis brain is composed of tissue-resident CD8+ T lymphocytes and B cells. *Brain.* 2018;141(7):2066-2082.
24. Frischer JM, Bramow S, Dal-Bianco A, et al. The relation between inflammation and neurodegeneration in multiple sclerosis brains. *Brain.* 2009;132(Pt 5):1175-1189.
25. Mews I, Bergmann M, Bunkowski S, Gullotta F, Brück W. Oligodendrocyte and axon pathology in clinically silent multiple sclerosis lesions. *Mult Scler.* 1998;4(2):55-62.
26. Sofroniew MV, Vinters HV. Astrocytes: biology and pathology. *Acta Neuropathol.* 2010;119(1):7-35.
27. Filippi M, Preziosa P, Banwell BL, et al. Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines. *Brain.* 2019;142(7):1858-1875.
28. Sander M, Frankfurt M. Hirnrindenbefunde bei multipler Sklerose. *Monatschrift Psychiatrie Neurol.* 1898;4:427-436.
29. Geurts JJ, Barkhof F. Grey matter pathology in multiple sclerosis. *Lancet Neurol.* 2008;7(9):841-851.
30. Peterson JW, Bo L, Mork S, Chang A, Trapp BD. Transected neurites, apoptotic neurons, and reduced inflammation in cortical multiple sclerosis lesions. *Ann Neurol.* 2001;50(3):389-400.
31. Bo L, Vedeler CA, Nyland HI, Trapp BD, Mork SJ. Subpial demyelination in the cerebral cortex of multiple sclerosis patients. *J Neuropathol Exp Neurol.* 2003;62(7):723-732.
32. Bø L, Vedeler CA, Nyland H, Trapp BD, Mørk SJ. Intracortical multiple sclerosis lesions are not associated with increased lymphocyte infiltration. *Mult Scler.* 2003;9(4):323-331.
33. Brink BP, Veerhuis R, Breij EC, van der Valk P, Dijkstra CD, Bö L. The pathology of multiple sclerosis is location-dependent: no significant

complement activation is detected in purely cortical lesions. *J Neuropathol Exp Neurol.* 2005;64(2):147-155.
34. van Horssen J, Brink BP, de Vries HE, van der Valk P, Bø L. The blood-brain barrier in cortical multiple sclerosis lesions. *J Neuropathol Exp Neurol.* 2007;66(4):321-328.
35. Serafini B, Rosicarelli B, Magliozzi R, Stigliano E, Aloisi F. Detection of ectopic B-cell follicles with germinal centers in the meninges of patients with secondary progressive multiple sclerosis. *Brain Pathol.* 2004;14(2):164-174.
36. Magliozzi R, Howell OW, Reeves C, et al. A Gradient of neuronal loss and meningeal inflammation in multiple sclerosis. *Ann Neurol.* 2010;68(4):477-493.
37. Allen IV, McKeown SR. A histological, histochemical and biochemical study of the macroscopically normal white matter in multiple sclerosis. *J Neurol Sci.* 1979;41(1):81-91.
38. Allen IV, McQuaid S, Mirakhur M, Nevin G. Pathological abnormalities in the normal-appearing white matter in multiple sclerosis. *Neurol Sci.* 2001;22(2):141-144.
39. Dziedzic T, Metz I, Dallenga T, et al. Wallerian degeneration: a major component of early axonal pathology in multiple sclerosis. *Brain Pathol.* 2010;20(5):976-985.
40. Klaver R, Popescu V, Voorn P, et al. Neuronal and Axonal Loss in Normal-Appearing Gray Matter and Subpial Lesions in Multiple Sclerosis. *J Neuropathol Exp Neurol.* 2015;74(5):453-458.
41. Cifelli A, Arridge M, Jezzard P, Esiri MM, Palace J, Matthews PM. Thalamic neurodegeneration in multiple sclerosis. *Ann Neurol.* 2002;52(5):650-653.
42. Haider L, Simeonidou C, Steinberger G, et al. Multiple sclerosis deep grey matter: the relation between demyelination, neurodegeneration, inflammation and iron. *J Neurol Neurosurg Psych.* 2014;85(12):1386-1395.
43. Pitt D, Werner P, Raine CS. Glutamate excitotoxicity in a model of multiple sclerosis. *Nat Med.* 2000;6(1):67-70.
44. Fischer MT, Sharma R, Lim JL, et al. NADPH oxidase expression in active multiple sclerosis lesions in relation to oxidative tissue damage and mitochondrial injury. *Brain.* 2012;135(Pt 3):886-899.
45. Haider L, Fischer MT, Frischer JM, et al. Oxidative damage in multiple sclerosis lesions. *Brain.* 2011;134(Pt 7):1914-1924.
46. Dutta R, McDonough J, Yin X, et al. Mitochondrial dysfunction as a cause of axonal degeneration in multiple sclerosis patients. *Ann Neurol.* 2006;59(3):478-489.
47. Craner MJ, Newcombe J, Black JA, Hartle C, Cuzner ML, Waxman SG. Molecular changes in neurons in multiple sclerosis: altered axonal expression of Nav1.2 and Nav1.6 sodium channels and Na+/Ca2+ exchanger. *Proc Natl Acad Sci U S A.* 2004;101(21):8168-8173.
48. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. *Nat Rev Immunol.* 2015;15(9):545-558.
49. Stys PK. Axonal degeneration in multiple sclerosis: is it time for neuroprotective strategies? *Ann Neurol.* 2004;55(5):601-603.

50. 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;11:23.
51. 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):99-106.
52. Eshaghi A, Marinescu RV, Young AL, et al. Progression of regional grey matter atrophy in multiple sclerosis. *Brain.* 2018;141(6):1665-1677.
53. McGinley MP, Goldschmidt CH, Rae-Grant AD. Diagnosis and Treatment of Multiple Sclerosis: A Review. *JAMA.* 2021;325(8):765-779.
54. Racosta JM, Kimpinski K, Morrow SA, Kremenchutzky M. Autonomic dysfunction in multiple sclerosis. *Auton Neurosci.* 2015;193:1-6.
55. Chiaravalloti ND, DeLuca J. Cognitive impairment in multiple sclerosis. *Lancet Neurol.* 2008;7(12):1139-1151.
56. Di Filippo M, Portaccio E, Mancini A, Calabresi P. Multiple sclerosis and cognition: synaptic failure and network dysfunction. *Nat Rev Neurosci.* 2018;19(10):599-609.
57. Kurtzke JF. Rating neurologic impairment in multiple sclerosis. *Neurology.* 1983;33(11):1444.
58. Polman CH, Rudick RA. The Multiple Sclerosis Functional Composite. A clinically meaningful measure of disability. *Neurology.* 2010;74(17 Supplement 3):S8-S15.
59. Oreja-Guevara C, Ayuso Blanco T, Brieva Ruiz L, Hernández Pérez MÁ, Meca-Lallana V, Ramió-Torrentà L. Cognitive Dysfunctions and Assessments in Multiple Sclerosis. *Front Neurol.* 2019;10.
60. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* 2018;17(2):162-173.
61. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology.* 2014;83(3):278-286.
62. Lassmann H. Pathogenic Mechanisms Associated With Different Clinical Courses of Multiple Sclerosis. *Front Immunol.* 2019;9:3116.
63. Filippi M, Rocca MA, Calabrese M, et al. Intracortical lesions. *Neurology.* 2010;75(22):1988.
64. Dalton CM, Chard DT, Davies GR, 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):1101-1107.
65. Rocca MA. clinically isolated syndrome suggestive of Multiple sclerosis: Dynamic Patterns of Gray and White Matter Changes—A 2-year MR Imaging Study. *Radiology.* 2015;278(3):841-853.
66. Kappos L, Moeri D, Radue EW, et al. Predictive value of gadolinium-enhanced magnetic resonance imaging for relapse rate and changes in disability or impairment in multiple sclerosis: a meta-analysis. Gadolinium MRI Meta-analysis Group. *Lancet.* 1999;353(9157):964-969.

67. Vrenken H, Pouwels PJW, Geurts JJG, et al. Altered diffusion tensor in multiple sclerosis normal-appearing brain tissue: Cortical diffusion changes seem related to clinical deterioration. *J Magn Reson Imaging.* 2006;23(5):628-636.
68. Calabrese M, Agosta F, Rinaldi F, et al. Cortical lesions and atrophy associated with cognitive impairment in relapsing-remitting multiple sclerosis. *Arch Neurol.* 2009;66(9):1144-1150.
69. Absinta M, Sati P, Masuzzo F, et al. Association of Chronic Active Multiple Sclerosis Lesions With Disability In Vivo. *JAMA Neurol.* 2019;76(12):1474-1483.
70. Lassmann H, Brück W, Lucchinetti CF. The Immunopathology of Multiple Sclerosis: An Overview. *Brain Pathol.* 2007;17(2):210-218.
71. Hauser SL, Cree BAC. Treatment of Multiple Sclerosis: A Review. *Am J Med.* 2020;133(12):1380-1390.e1382.
72. Kappos L, Bar-Or A, Cree BAC, et al. Siponimod versus placebo in secondary progressive multiple sclerosis (EXPAND): a double-blind, randomised, phase 3 study. *Lancet.* 2018;391(10127):1263-1273.
73. Montalban X, Hauser SL, Kappos L, et al. Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis. *N Engl J Med.* 2016;376(3):209-220.
74. Ontaneda D, Thompson AJ, Fox RJ, Cohen JA. Progressive multiple sclerosis: prospects for disease therapy, repair, and restoration of function. *Lancet.* 2017;389(10076):1357-1366.
75. Chataway J, De Angelis F, Connick P, et al. Efficacy of three neuroprotective drugs in secondary progressive multiple sclerosis (MS-SMART): a phase 2b, multiarm, double-blind, randomised placebo-controlled trial. *Lancet Neurol.* 2020;19(3):214-225.
76. McKay KA, Tremlett H, Fisk JD, et al. Psychiatric comorbidity is associated with disability progression in multiple sclerosis. *Neurology.* 2018;90(15):e1316.
77. Marrie RA, Horwitz R, Cutter G, Tyry T, Campagnolo D, Vollmer T. The burden of mental comorbidity in multiple sclerosis: frequent, underdiagnosed, and undertreated. *Mult Scler.* 2009;15(3):385-392.
78. Marrie RA, Rudick R, Horwitz R, et al. Vascular comorbidity is associated with more rapid disability progression in multiple sclerosis. *Neurology.* 2010;74(13):1041.
79. Geraldes R, Esiri MM, DeLuca GC, Palace J. Age-related small vessel disease: a potential contributor to neurodegeneration in multiple sclerosis. *Brain Pathol.* 2017;27(6):707-722.
80. Ascherio A, Munger KL, White R, et al. Vitamin D as an early predictor of multiple sclerosis activity and progression. *JAMA Neurol.* 2014;71(3):306-314.
81. Runia TF, Hop WC, de Rijke YB, Buljevac D, Hintzen RQ. Lower serum vitamin D levels are associated with a higher relapse risk in multiple sclerosis. *Neurology.* 2012;79(3):261-266.
82. Fitzgerald KC, Salter A, Tyry T, Fox RJ, Cutter G, Marrie RA. Measures of general and abdominal obesity and disability severity in a large population of people with multiple sclerosis. *Mult Scler.* 2020;26(8):976-986.

83. Tettey P, Simpson S, Taylor B, et al. An adverse lipid profile and increased levels of adiposity significantly predict clinical course after a first demyelinating event. *J Neurol Neurosurg Psychiatry.* 2017;88(5):395-401.
84. Zivadinov R, Weinstock-Guttman B, Hashmi K, et al. Smoking is associated with increased lesion volumes and brain atrophy in multiple sclerosis. *Neurology.* 2009;73(7):504-510.
85. Horakova D, Zivadinov R, Weinstock-Guttman B, et al. Environmental Factors Associated with Disease Progression after the First Demyelinating Event: Results from the Multi-Center SET Study (P05.134). *Neurology.* 2013;80(7 Supplement):P05.134.
86. Healy BC, Ali EN, Guttmann CR, et al. Smoking and disease progression in multiple sclerosis. *Arch Neurol.* 2009;66(7):858-864.
87. Munger KL, Fitzgerald KC, Freedman MS, et al. No association of multiple sclerosis activity and progression with EBV or tobacco use in BENEFIT. *Neurology.* 2015;85(19):1694-1701.
88. Kvistad S, Myhr KM, Holmøy T, et al. No association of tobacco use and disease activity in multiple sclerosis. *Neurol Neuroimmunol Neuroinflamm.* 2016;3(4):e260.
89. Vine MF, Hulka BS, Margolin BH, et al. Cotinine concentrations in semen, urine, and blood of smokers and nonsmokers. *Am J Public Health.* 1993;83(9):1335-1338.
90. Connor Gorber S, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. *Nicotine Tob Res.* 2009;11(1):12-24.
91. Hedström AK, Hillert J, Olsson T, Alfredsson L. Nicotine might have a protective effect in the etiology of multiple sclerosis. *Mult Scler.* 2013;19(8):1009-1013.
92. Hedström AK, Bäärnhielm M, Olsson T, Alfredsson L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. *Neurology.* 2009;73(9):696-701.
93. Hedström AK, Olsson T, Alfredsson L. Cotinine as a measure of smoking in observational studies of multiple sclerosis. *Mult Scler.* 2020;27(8):1293-1296.
94. Petersen ER, Oturai AB, Koch-Henriksen N, et al. Smoking affects the interferon beta treatment response in multiple sclerosis. *Neurology.* 2018;90(7):e593-e600.
95. Petersen ER, Søndergaard HB, Laursen JH, et al. Smoking is associated with increased disease activity during natalizumab treatment in multiple sclerosis. *Mult Scler.* 2018;25(9):1298-1305.
96. Pittas F, Ponsonby A-L, van der Mei IAF, et al. Smoking is associated with progressive disease course and increased progression in clinical disability in a prospective cohort of people with multiple sclerosis. *J Neurol.* 2009;256(4):577.
97. Weiland TJ, Hadgkiss EJ, Jelinek GA, Pereira NG, Marck CH, van der Meer DM. The association of alcohol consumption and smoking with quality of life, disability and disease activity in an international sample of people with multiple sclerosis. *J Neurol Sci.* 2014;336(1-2):211-219.

98. Correale J, Farez MF. Smoking worsens multiple sclerosis prognosis: Two different pathways are involved. *J Neuroimmunol.* 2015;281:23-34.
99. Ramanujam R, Hedström AK, Manouchehrinia A, et al. Effect of Smoking Cessation on Multiple Sclerosis Prognosis. *JAMA Neurol.* 2015;72(10):1117-1123.
100. Hernán MA, Jick SS, Logroscino G, Olek MJ, Ascherio A, Jick H. Cigarette smoking and the progression of multiple sclerosis. *Brain.* 2005;128(Pt 6):1461-1465.
101. Degelman ML, Herman KM. Smoking and multiple sclerosis: A systematic review and meta-analysis using the Bradford Hill criteria for causation. *Mult Scler Relat Disord.* 2017;17:207-216.
102. Handel AE, Williamson AJ, Disanto G, Dobson R, Giovannoni G, Ramagopalan SV. Smoking and multiple sclerosis: an updated meta-analysis. *PloS one.* 2011;6(1):e16149.
103. Manouchehrinia A, Tench CR, Maxted J, Bibani RH, Britton J, Constantinescu CS. Tobacco smoking and disability progression in multiple sclerosis: United Kingdom cohort study. *Brain.* 2013;136(7):2298-2304.
104. Wesnes K, Myhr KM, Riise T, et al. Low vitamin D, but not tobacco use or high BMI, is associated with long-term disability progression in multiple sclerosis. *Mult Scler Relat Disord.* 2021;50:102801.
105. Koch M, van Harten A, Uyttenboogaart M, De Keyser J. Cigarette smoking and progression in multiple sclerosis. *Neurology.* 2007;69(15):1515.
106. Bermel RA, Bakshi R. The measurement and clinical relevance of brain atrophy in multiple sclerosis. *Lancet Neurol.* 2006;5(2):158-170.
107. Kappus N, Weinstock-Guttman B, Hagemeier J, et al. Cardiovascular risk factors are associated with increased lesion burden and brain atrophy in multiple sclerosis. *J Neurol Neurosurg Psychiatry.* 2016;87(2):181-187.
108. Graetz C, Groger A, Luessi F, et al. Association of smoking but not HLA-DRB1*15:01, APOE or body mass index with brain atrophy in early multiple sclerosis. *Mult Scler.* 2018;25(5):661-668.
109. Rosso M, Chitnis T. Association Between Cigarette Smoking and Multiple Sclerosis: A Review. *JAMA Neurol.* 2020;77(2):245-253.
110. Wingerchuk DM. Smoking: effects on multiple sclerosis susceptibility and disease progression. *Ther Adv Neurol Disord.* 2011;5(1):13-22.
111. Hossain M, Sathe T, Fazio V, et al. Tobacco smoke: a critical etiological factor for vascular impairment at the blood-brain barrier. *Brain Res.* 2009;1287:192-205.
112. Dobson R, Rudick RA, Turner B, Schmierer K, Giovannoni G. Assessing treatment response to interferon-β: is there a role for MRI? *Neurology.* 2014;82(3):248-254.
113. Giovannoni G, Turner B, Gnanapavan S, Offiah C, Schmierer K, Marta M. Is it time to target no evident disease activity (NEDA) in multiple sclerosis? *Mult Scler Relat Disord.* 2015;4(4):329-333.
114. Chard D, Trip SA. Resolving the clinico-radiological paradox in multiple sclerosis. *F1000Res.* 2017;6:1828.

115. Saade C, Bou-Fakhredin R, Yousem DM, Asmar K, Naffaa L, El-Merhi F. Gadolinium and Multiple Sclerosis: Vessels, Barriers of the Brain, and Glymphatics. *AJNR Am J Neuroradiol.* 2018;39(12):2168-2176.
116. Geurts JJ, Bö L, Pouwels PJ, Castelijns JA, Polman CH, Barkhof F. Cortical lesions in multiple sclerosis: combined postmortem MR imaging and histopathology. *AJNR Am J Neuroradiol.* 2005;26(3):572-577.
117. Barkhof F. The clinico-radiological paradox in multiple sclerosis revisited. *Curr Opin Neurol.* 2002;15(3):239-245.
118. Meyer-Moock S, Feng YS, Maeurer M, Dippel FW, Kohlmann T. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. *BMC Neurol.* 2014;14:58.
119. Rocca MA, Comi G, Filippi M. The Role of T1-Weighted Derived Measures of Neurodegeneration for Assessing Disability Progression in Multiple Sclerosis. *Front Neurol.* 2017;8:433.
120. Calabrese M, Rocca MA, Atzori M, et al. A 3-year magnetic resonance imaging study of cortical lesions in relapse-onset multiple sclerosis. *Ann Neurol.* 2010;67(3):376-383.
121. Brown JWL, Chard DT. The role of MRI in the evaluation of secondary progressive multiple sclerosis. *Expert Rev Neurother.* 2016;16(2):157-171.
122. Popescu V, Klaver R, Voorn P, et al. What drives MRI-measured cortical atrophy in multiple sclerosis? *Mult Scler.* 2015;21(10):1280-1290.
123. Bergsland N, Horakova D, Dwyer MG, 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.* 2012;33(8):1573-1578.
124. 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.
125. Fisher E, Lee J-C, Nakamura K, Rudick RA. Gray matter atrophy in multiple sclerosis: A longitudinal study. *Ann Neurol.* 2008;64(3):255-265.
126. Steenwijk MD, Geurts JJ, Daams M, et al. Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. *Brain.* 2016;139(Pt 1):115-126.
127. Pagani E, Rocca MA, Gallo A, et al. Regional brain atrophy evolves differently in patients with multiple sclerosis according to clinical phenotype. *AJNR Am J Neuroradiol.* 2005;26(2):341-346.
128. 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.
129. 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.
130. 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.

131. 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.
132. 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.
133. Sepulcre J, Masdeu JC, Goñi J, et al. Fatigue in multiple sclerosis is associated with the disruption of frontal and parietal pathways. *Mult Scler.* 2009;15(3):337-344.
134. Pravata E, Rocca MA, Valsasina P, et al. Gray matter trophism, cognitive impairment, and depression in patients with multiple sclerosis. *Mult Scler.* 2017;23(14):1864-1874.
135. 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.
136. Sastre-Garriga J, Pareto D, Battaglini M, et al. MAGNIMS consensus recommendations on the use of brain and spinal cord atrophy measures in clinical practice. *Nat Rev Neurol.* 2020;16(3):171-182.
137. 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.
138. Rovira À, Wattjes MP, Tintoré M, et al. MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis—clinical implementation in the diagnostic process. *Nat Rev Neurol.* 2015;11(8):471-482.
139. Hannoun S, Baalbaki M, Haddad R, et al. Gadolinium effect on thalamus and whole brain tissue segmentation. *Neuroradiology.* 2018;60(11):1167-1173.
140. Warntjes JB, Tisell A, Landtblom AM, Lundberg P. Effects of gadolinium contrast agent administration on automatic brain tissue classification of patients with multiple sclerosis. *AJNR Am J Neuroradiol.* 2014;35(7):1330-1336.
141. Roberto CA, Mayer LE, Brickman AM, et al. Brain tissue volume changes following weight gain in adults with anorexia nervosa. *Int J Eat Disord.* 2011;44(5):406-411.
142. Janowitz D, Wittfeld K, Terock J, et al. Association between waist circumference and gray matter volume in 2344 individuals from two adult community-based samples. *Neuroimage.* 2015;122:149-157.
143. Duning T, Kloska S, Steinsträter O, Kugel H, Heindel W, Knecht S. Dehydration confounds the assessment of brain atrophy. *Neurology.* 2005;64(3):548-550.
144. Nakamura K, Brown RA, Narayanan S, Collins DL, Arnold DL. Diurnal fluctuations in brain volume: Statistical analyses of MRI from large populations. *Neuroimage.* 2015;118:126-132.
145. Jakimovski D, Gandhi S, Paunkoski I, et al. Hypertension and heart disease are associated with development of brain atrophy in multiple sclerosis: a 5-year longitudinal study. *Eur J Neurol.* 2018;26(1):87-e8.
146. Friedman JI, Tang CY, de Haas HJ, et al. Brain Imaging Changes Associated With Risk Factors for Cardiovascular and Cerebrovascular Disease in Asymptomatic Patients. *JACC Cardiovasc Imaging.* 2014;7(10):1039-1053.

147. 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):532-538.
148. 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):1175-1181.
149. Favaretto A, Lazzarotto A, Margoni M, Poggiali D, Gallo P. Effects of disease modifying therapies on brain and grey matter atrophy in relapsing remitting multiple sclerosis. *Mult Scler Demyelinating Disord.* 2018;3(1):1.
150. Sotirchos ES, Gonzalez-Caldito N, Dewey BE, et al. Effect of disease-modifying therapies on subcortical gray matter atrophy in multiple sclerosis. *Mult Scler.* 2020;26(3):312-321.
151. Traboulsee A, Pelletier D, Comi G, et al. Alemtuzumab reduces the rate of brain volume loss in RRMS patients who switched from SC IFNB-1a to alemtuzumab (4-year follow-up of the CARE-MS I and II Studies). Paper presented at: MULTIPLE SCLEROSIS JOURNAL2016.
152. Comabella M, Montalban X. Body fluid biomarkers in multiple sclerosis. *Lancet Neurol.* 2014;13(1):113-126.
153. Yuan A, Rao MV, Veeranna, Nixon RA. Neurofilaments and Neurofilament Proteins in Health and Disease. *Cold Spring Harb Perspect Biol.* 2017;9(4).
154. Gaetani L, Blennow K, Calabresi P, Di Filippo M, Parnetti L, Zetterberg H. Neurofilament light chain as a biomarker in neurological disorders. *J Neurol Neurosurg Psychiatry.* 2019;90(8):870-881.
155. Reiber H. Proteins in cerebrospinal fluid and blood: barriers, CSF flow rate and source-related dynamics. *Restor Neurol Neurosci.* 2003;21(3-4):79-96.
156. Disanto G, Barro C, Benkert P, et al. Serum Neurofilament light: A biomarker of neuronal damage in multiple sclerosis. *Ann Neurol.* 2017;81(6):857-870.
157. Rosengren LE, Karlsson JE, Karlsson JO, Persson LI, Wikkelso C. Patients with amyotrophic lateral sclerosis and other neurodegenerative diseases have increased levels of neurofilament protein in CSF. *J Neurochem.* 1996;67(5):2013-2018.
158. Kuhle J, Barro C, Andreasson U, et al. Comparison of three analytical platforms for quantification of the neurofilament light chain in blood samples: ELISA, electrochemiluminescence immunoassay and Simoa. *Clin Chem Lab Med.* 2016;54(10):1655-1661.
159. Gisslén M, Price RW, Andreasson U, et al. Plasma Concentration of the Neurofilament Light Protein (NFL) is a Biomarker of CNS Injury in HIV Infection: A Cross-Sectional Study. *EBioMedicine.* 2016;3:135-140.
160. Wilson DH, Rissin DM, Kan CW, et al. The Simoa HD-1 Analyzer: A Novel Fully Automated Digital Immunoassay Analyzer with Single-Molecule Sensitivity and Multiplexing. *J Lab Autom.* 2016;21(4):533-547.
161. Kuhle J, Barro C, Hrusovsky K, et al. International multi-site analytical validation of the Simoa NF-light assay in human serum samples from multiple sclerosis patients. Paper presented at: Multiple Sclerosis Journal2018.

162. Altmann P, Leutmezer F, Zach H, et al. Serum neurofilament light chain withstands delayed freezing and repeated thawing. *Sci Rep.* 2020;10(1):19982.
163. Keshavan A, Heslegrave A, Zetterberg H, Schott JM. Stability of blood-based biomarkers of Alzheimer's disease over multiple freeze-thaw cycles. *Alzheimers Dement (Amst).* 2018;10:448-451.
164. Bittner S, Steffen F, Uphaus T, et al. Clinical implications of serum neurofilament in newly diagnosed MS patients: A longitudinal multicentre cohort study. *EBioMedicine.* 2020;56:102807.
165. Barro C, Benkert P, Disanto G, et al. Serum neurofilament as a predictor of disease worsening and brain and spinal cord atrophy in multiple sclerosis. *Brain.* 2018;141(8):2382-2391.
166. Varhaug KN, Barro C, Bjornevik K, et al. Neurofilament light chain predicts disease activity in relapsing-remitting MS. *Neurol Neuroimmunol Neuroinflamm.* 2018;5(1):e422.
167. Srpova B, Uher T, Hrnciarova T, et al. Serum neurofilament light chain reflects inflammation-driven neurodegeneration and predicts delayed brain volume loss in early stage of multiple sclerosis. *Mult Scler.* 2021;27(1):52-60.
168. Bittner S, Oh J, Havrdová EK, Tintoré M, Zipp F. The potential of serum neurofilament as biomarker for multiple sclerosis. *Brain.* 2021;144(10):2954-2963.
169. Rosso M, Gonzalez CT, Healy BC, et al. Temporal association of sNfL and gad-enhancing lesions in multiple sclerosis. *Ann Clin Transl Neurol.* 2020;7(6):945-955.
170. Sejbaek T, Nielsen HH, Penner N, et al. Dimethyl fumarate decreases neurofilament light chain in CSF and blood of treatment naïve relapsing MS patients. *J Neurol Neurosurg Psychiatry.* 2019;90(12):1324-1330.
171. Delcoigne B, Manouchehrinia A, Barro C, et al. Blood neurofilament light levels segregate treatment effects in multiple sclerosis. *Neurology.* 2020;94(11):e1201-e1212.
172. Jakimovski D, Kuhle J, Ramanathan M, et al. Serum neurofilament light chain levels associations with gray matter pathology: a 5-year longitudinal study. *Ann Clin Transl Neurol.* 2019;6(9):1757-1770.
173. Bhan A, Jacobsen C, Myhr KM, Dalen I, Lode K, Farbu E. Neurofilaments and 10-year follow-up in multiple sclerosis. *Mult Scler.* 2018;24(10):1301-1307.
174. Manouchehrinia A, Stridh P, Khademi M, et al. Plasma neurofilament light levels are associated with risk of disability in multiple sclerosis. *Neurology.* 2020;94(23):e2457.
175. Thebault S, Abdoli M, Fereshtehnejad S-M, Tessier D, Tabard-Cossa V, Freedman MS. Serum neurofilament light chain predicts long term clinical outcomes in multiple sclerosis. *Sci Rep.* 2020;10(1):10381.
176. Cantó E, Barro C, Zhao C, et al. Association Between Serum Neurofilament Light Chain Levels and Long-term Disease Course Among Patients With Multiple Sclerosis Followed up for 12 Years. *JAMA Neurol.* 2019;76(11):1359-1366.

177. Chitnis T, Gonzalez C, Healy BC, et al. Neurofilament light chain serum levels correlate with 10-year MRI outcomes in multiple sclerosis. *Ann Clin Transl Neurol.* 2018;5(12):1478-1491.
178. Tsagkas C, Parmar K, Pezold S, et al. Classification of multiple sclerosis based on patterns of CNS regional atrophy covariance. *Hum Brain Mapp.* 2021;42(8):2399-2415.
179. Filippi P, Vestenická V, Siarnik P, et al. Neurofilament light chain and MRI volume parameters as markers of neurodegeneration in multiple sclerosis. *Neuro Endocrinol Lett.* 2020;41(1):17-26.
180. Khalil M, Pirpamer L, Hofer E, et al. Serum neurofilament light levels in normal aging and their association with morphologic brain changes. *Nat Commun.* 2020;11(1):812.
181. Manouchehrinia A, Piehl F, Hillert J, et al. Confounding effect of blood volume and body mass index on blood neurofilament light chain levels. *Ann Clin Transl Neurol.* 2020;7(1):139-143.
182. Lewczuk P, Ermann N, Andreasson U, et al. Plasma neurofilament light as a potential biomarker of neurodegeneration in Alzheimer's disease. *Alzheimers Res Ther.* 2018;10(1):71.
183. Gille B, De Schaepdryver M, Goossens J, et al. Serum neurofilament light chain levels as a marker of upper motor neuron degeneration in patients with Amyotrophic Lateral Sclerosis. *Neuropathol Appl Neurobiol.* 2019;45(3):291-304.
184. Hansson O, Janelidze S, Hall S, et al. Blood-based NfL: A biomarker for differential diagnosis of parkinsonian disorder. *Neurology.* 2017;88(10):930-937.
185. Tiedt S, Duering M, Barro C, et al. Serum neurofilament light: A biomarker of neuroaxonal injury after ischemic stroke. *Neurology.* 2018;91(14):e1338-e1347.
186. Altmann P, De Simoni D, Kaider A, et al. Increased serum neurofilament light chain concentration indicates poor outcome in Guillain-Barré syndrome. *J Neuroinflamm.* 2020;17(1):86.
187. Shahim P, Zetterberg H, Tegner Y, Blennow K. Serum neurofilament light as a biomarker for mild traumatic brain injury in contact sports. *Neurology.* 2017;88(19):1788-1794.
188. Korley FK, Goldstick J, Mastali M, et al. Serum NfL (Neurofilament Light Chain) Levels and Incident Stroke in Adults With Diabetes Mellitus. *Stroke.* 2019;50(7):1669-1675.
189. Gattringer T, Pinter D, Enzinger C, et al. Serum neurofilament light is sensitive to active cerebral small vessel disease. *Neurology.* 2017;89(20):2108-2114.
190. Boehnke SE, Robertson EL, Armitage-Brown B, et al. The effect of lumbar puncture on the neurodegeneration biomarker neurofilament light in macaque monkeys. *Alzheimers Dement (Amst).* 2020;12(1):e12069.
191. Benkert P, Meier S, Schaedelin S, et al. Serum neurofilament light chain for individual prognostication of disease activity in people with multiple sclerosis: a retrospective modelling and validation study. *Lancet Neurol.* 2022;21(3):246-257.

192. Barro C, Chitnis T, Weiner HL. Blood neurofilament light: a critical review of its application to neurologic disease. *Ann Clin Transl Neurol.* 2020;7(12):2508-2523.
193. Axelsson M, Malmeström C, Nilsson S, Haghighi S, Rosengren L, Lycke J. Glial fibrillary acidic protein: a potential biomarker for progression in multiple sclerosis. *J Neurol.* 2011;258(5):882-888.
194. Norgren N, Sundström P, Svenningsson A, Rosengren L, Stigbrand T, Gunnarsson M. Neurofilament and glial fibrillary acidic protein in multiple sclerosis. *Neurology.* 2004;63(9):1586-1590.
195. Högel H, Rissanen E, Barro C, et al. Serum glial fibrillary acidic protein correlates with multiple sclerosis disease severity. *Mult Scler.* 2018;26(2):210-219.
196. Liu JS, Zhao ML, Brosnan CF, Lee SC. Expression of inducible nitric oxide synthase and nitrotyrosine in multiple sclerosis lesions. *Am J Pathol.* 2001;158(6):2057-2066.
197. Bö L, Dawson TM, Wesselingh S, et al. Induction of nitric oxide synthase in demyelinating regions of multiple sclerosis brains. *Ann Neurol.* 1994;36(5):778-786.
198. Rejdak K, Eikelenboom MJ, Petzold A, et al. CSF nitric oxide metabolites are associated with activity and progression of multiple sclerosis. *Neurology.* 2004;63(8):1439.
199. Brundin L, Morcos E, Olsson T, Wiklund NP, Andersson M. Increased intrathecal nitric oxide formation in multiple sclerosis; cerebrospinal fluid nitrite as activity marker. *Eur J Neurol.* 1999;6(5):585-590.
200. Engel S, Friedrich M, Muthuraman M, et al. Intrathecal B-cell accumulation and axonal damage distinguish MRI-based benign from aggressive onset in MS. *Neurol Neuroimmunol Neuroinflamm.* 2019;6(5):e595.
201. 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.
202. Torkildsen O, Wergeland S, Bakke S, et al. omega-3 fatty acid treatment in multiple sclerosis (OFAMS Study): a randomized, double-blind, placebo-controlled trial. *Arch Neurol.* 2012;69(8):1044-1051.
203. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol.* 2001;50(1):121-127.
204. Biochemical verification of tobacco use and cessation. *Nicotine Tob Res.* 2002;4(2):149-159.
205. Ambrose JA, Barua RS. The pathophysiology of cigarette smoking and cardiovascular disease: an update. *J Am Coll Cardiol.* 2004;43(10):1731-1737.
206. Schmidt P, Gaser C, Arsic M, et al. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *NeuroImage.* 2012;59(4):3774-3783.
207. Battaglini M, Jenkinson M, De Stefano N. Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Hum Brain Mapp.* 2012;33(9):2062-2071.

208. Fischl B. FreeSurfer. *NeuroImage.* 2012;62(2):774-781.
209. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage.* 1999;9(2):179-194.
210. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage.* 2006;31(3):968-980.
211. Giavarina D. Understanding Bland Altman analysis. *Biochem Med.* 2015;25(2):141-151.
212. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol.* 1995;57(1):289-300.
213. Ciccarelli O, Barkhof F, Bodini B, et al. Pathogenesis of multiple sclerosis: insights from molecular and metabolic imaging. *Lancet Neurol.* 2014;13(8):807-822.
214. Correale J, Gaitán MI, Ysrraelit MC, Fiol MP. Progressive multiple sclerosis: from pathogenic mechanisms to treatment. *Brain.* 2017;140(3):527-546.
215. Geurts JJ, Stys PK, Minagar A, Amor S, Zivadinov R. Gray matter pathology in (chronic) MS: modern views on an early observation. *J Neurol Sci.* 2009;282(1-2):12-20.
216. Geurts JJ, Calabrese M, Fisher E, Rudick RA. Measurement and clinical effect of grey matter pathology in multiple sclerosis. *Lancet Neurol.* 2012;11(12):1082-1092.
217. Andravizou A, Dardiotis E, Artemiadis A, et al. Brain atrophy in multiple sclerosis: mechanisms, clinical relevance and treatment options. *Auto Immun Highlights.* 2019;10(1):7.
218. Filippi M, Rocca MA. MR Imaging of Gray Matter Involvement in Multiple Sclerosis: Implications for Understanding Disease Pathophysiology and Monitoring Treatment Efficacy. *AJNR Am J Neuroradiol.* 2010;31(7):1171.
219. Ellwardt E, Zipp F. Molecular mechanisms linking neuroinflammation and neurodegeneration in MS. *Exp Neurol.* 2014;262:8-17.
220. Hasan KM, Kamali A, Kramer LA. Mapping the human brain white matter tracts relative to cortical and deep gray matter using diffusion tensor imaging at high spatial resolution. *Magn Reson Imaging.* 2009;27(5):631-636.
221. Henry RG, Shieh M, Amirbekian B, Chung S, Okuda DT, Pelletier D. Connecting white matter injury and thalamic atrophy in clinically isolated syndromes. *J Neurol Sci.* 2009;282(1-2):61-66.
222. Bergsland N, Lagana MM, Tavazzi E, et al. Corticospinal tract integrity is related to primary motor cortex thinning in relapsing-remitting multiple sclerosis. *Mult Scler.* 2015;21(14):1771-1780.
223. Lokhande H, Rosso M, Tauhid S, et al. Serum NfL levels in the first five years predict 10-year thalamic fraction in patients with MS. *Mult Scler J Exp Transl Clin.* 2022;8(1):20552173211069348.
224. Neacsu V, Jasperse B, Korteweg T, et al. Agreement between different input image types in brain atrophy measurement in multiple sclerosis using SIENAX and SIENA. *J Magn Reson Imaging.* 2008;28(3):559-565.

225. Liu Y, Lukas C, Steenwijk MD, 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.* 2016;37(4):749-754.
226. Teunissen CE, Khalil M. Neurofilaments as biomarkers in multiple sclerosis. *Mult Scler.* 2012;18(5):552-556.
227. Khalil M. Are neurofilaments valuable biomarkers for long-term disease prognostication in MS? *Mult Scler.* 2018;24(10):1270-1271.
228. Bridel C, Leurs CE, van Lierop ZY, et al. Serum Neurofilament Light Association With Progression in Natalizumab-Treated Patients With Relapsing-Remitting Multiple Sclerosis. *Neurology.* 2021;97(19):e1898-e1905.
229. Cortese M, Munger KL, Martínez-Lapiscina EH, et al. Vitamin D, smoking, EBV, and long-term cognitive performance in MS: 11-year follow-up of BENEFIT. *Neurology.* 2020;94(18):e1950-e1960.
230. Concato J, Shah N, Horwitz RI. Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs. *N Engl J Med.* 2000;342(25):1887-1892.
231. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet.* 2002;359(9302):248-252.
232. Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health.* 2004;58(8):635.
233. Hernán MA, Robins JM. *Causal inference: What if.* Boca Raton: Chapman & Hall/CRC; 2020.
234. Greenhalgh T. How to read a paper : getting your bearings (deciding what the paper is about). *BMJ.* 1997;315(7102):243-246.
235. Cartwright N. Are RCTs the Gold Standard? *BioSocieties.* 2007;2(1):11-20.
236. Concato J. Observational versus experimental studies: what's the evidence for a hierarchy? *NeuroRx.* 2004;1(3):341-347.
237. Yetley EA, MacFarlane AJ, Greene-Finestone LS, et al. Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group. *Am J Clin Nutr.* 2017;105(1):249s-285s.
238. Rothman KJ, Greenland S. Causation and Causal Inference in Epidemiology. *Am J Public Health.* 2005;95(S1):S144-S150.
239. Shreffler J, Huecker MR. Type I and Type II Errors and Statistical Power. In: *StatPearls.* Treasure Island (FL): StatPearls Publishing Copyright © 2022, StatPearls Publishing LLC.; 2022.
240. Song JW, Chung KC. Observational studies: cohort and case-control studies. *Plast Reconstr Surg.* 2010;126(6):2234-2242.

Articles I-IV

I

RESEARCH ARTICLE OPEN ACCESS

Relationship Between White Matter Lesions and Gray Matter Atrophy in Multiple Sclerosis

A Systematic Review

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

Correspondence
Dr. Lie
i.lie@amsterdamumc.nl

Neurology® 2022;98:e1562-e1573. doi:10.1212/WNL.0000000000200006

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 2 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.

From the Department of Radiology and Nuclear Medicine (I.A.L., M.M.W., R.M.M., P.J.W.P., F.B., H.V.), MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, Location VUmc; Department of Clinical Medicine (I.A.L., Ø.T., L.B., K.-M.M.), University of Bergen, Norway; Medical Library (M.A.E.M.), Amsterdam UMC, Vrije Universiteit Amsterdam, the Netherlands; Institutes of Neurology and Healthcare Engineering (F.B.), UCL London, UK; Neuro-SysMed, Department of Neurology (I.A.L., Ø.T., L.B., K.-M.M.), Haukeland University Hospital, Bergen, Norway; Norwegian Multiple Sclerosis Competence Centre, Department of Neurology (L.B.), Haukeland University Hospital, Bergen, Norway.

Go to Neurology.org/N for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

The article processing charge was funded by Vrije Universiteit Amsterdam.

Glossary

CGM = cortical GM; **CIS** = clinically isolated syndrome; **DGM** = deep GM; **GM** = gray matter; **LV** = lesion volume; **MS** = multiple sclerosis; **PPMS** = primary progressive MS; **RRMS** = relapsing-remitting MS; **SPMS** = secondary progressive MS; **WM** = white matter.

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, links.lww.com/WNL/B816. 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, links.lww.com/WNL/B816), 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.

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, links.lww.com/WNL/B816, 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]

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]

Figure 1 Flowchart Demonstrating the Selection Process

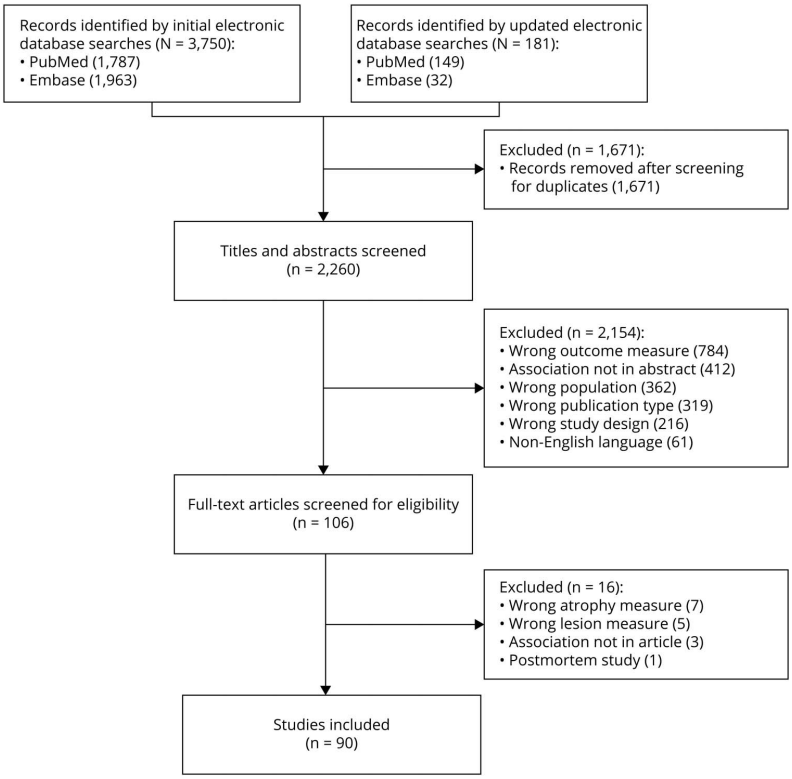


Table 1 Study Design in Included Studies

Study design	No.
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

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, links.lww.com/WNL/B816, and a more detailed discussion of results of each section is given in eAppendix 2.

Figure 2 illustrates the main results from this section.

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).

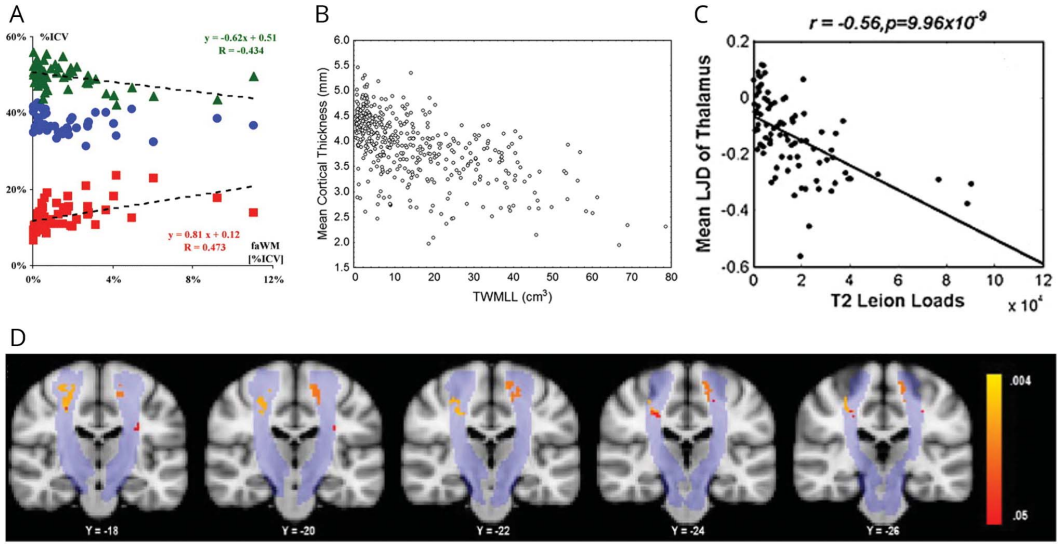
Three of 5 studies considering longitudinal WM lesion changes[e14,e24,e27-e29] with a follow-up period between 1 and 4 years observed significant associations between longitudinal changes in T1[e27] and T2[e24,e27,e28] LV and GM atrophy progression (p values ranging from 0.0004–0.03).

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

Figure 2 RRMS Shows Consistent Associations Between WM Lesions and GM Volume



(A): Scatterplot of the fractional volumes of gray matter (fGM, green triangles), white matter (fWM, blue circles), and CSF (fCSF, red boxes) vs the fractional volume of abnormal white matter (faWM), all expressed as percentages of intracranial volume. fGM and fCSF values are adjusted to patients mean age (35.6 years). When 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 white matter lesion volume) is not significantly changed with increasing faWM. Reproduced from Quarantelli et al., 2003,[e20] with permission from Elsevier. (B) Scatterplot showing the relationship between mean cortical thickness in millimeters and total white matter lesion load (TWMLL) in cubic centimeters in 425 patients with relapsing-remitting multiple sclerosis (RRMS). Reproduced from Charil et al., 2007,[e35] with permission from Elsevier. (C) Correlation of the mean logarithm of the jacobian determinant (LJD; a measure of atrophy) with T2 lesion load in 88 patients with RRMS for the thalamus. Reproduced from Tao et al., 2009.[e37] with permission from Elsevier. (D) Lesional voxels that significantly correlate with primary motor cortex thickness are shown in red-yellow. Probabilistic corticospinal tract atlas is shown in light blue. Reproduced from Bergsland et al., 2015,[e47] with permission from SAGE Publications.

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]

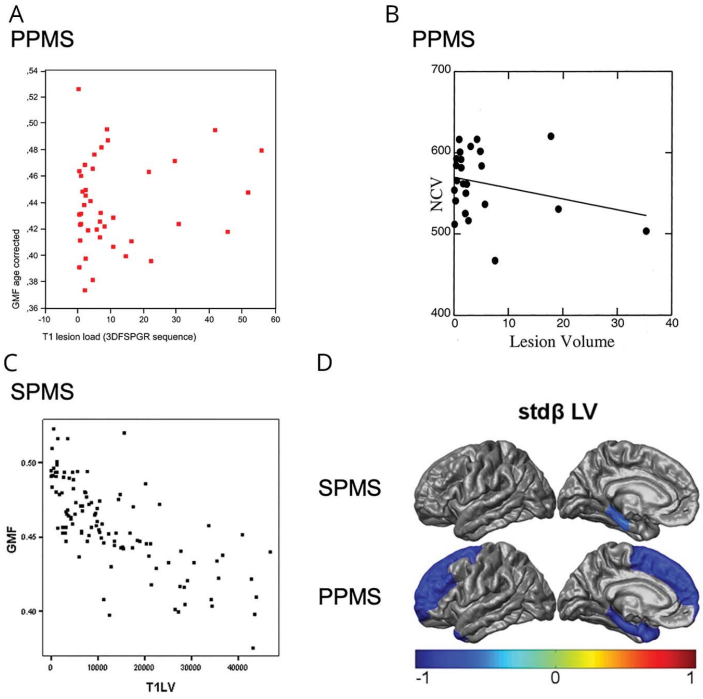
DGM in RRMS

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/

Figure 3 Progressive MS Shows Varying Associations Between WM Lesions and GM Volume



(A) Primary progressive multiple sclerosis (PPMS): In 43 patients with PPMS, gray matter (GM) fraction (GMF) corrected for age is plotted against T1 lesion load (in milliliters; derived from 3-dimensional fast spoiled gradient recalled echo scans), illustrating an absence of correlation. Reproduced from Sastre-Garriga et al., 2004,[e59] with permission from Elsevier. (B) PPMS: data illustrating the absence of correlation between normalized cortical volume and T2 lesion volume in 25 patients with PPMS (Spearman rank coefficient $r = -0.1$, $p = 0.6$). Reproduced from De Stefano et al., 2003,[e31] with permission from Wolters Kluwer Health, Inc. (C) Secondary progressive multiple sclerosis (SPMS): scatterplot of T1 lesion volume (T1LV) vs GMF in 117 patients with SPMS, illustrating a significant correlation ($r = -0.72$, $p < 0.001$). Reproduced from Furby et al., 2009,[e58] with permission from SAGE Publications. (D) Graphic visualization of the cross-sectional relationship between regional cortical thickness and white matter (WM) measures in 53 patients with long-standing 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, colors correspond to the standardized beta values of lesion volume in the connected tracts for the respective regional model. Reproduced from Steenwijk et al., 2015,[e46] with permission from John Wiley and Sons.

enlarging T2 lesions or new gadolinium-enhancing lesions ($p = 0.024$).[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, links.lww.com/WNL/B816, 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]

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, links.lww.com/WNL/B816, 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, links.lww.com/WNL/B816. 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, links.lww.com/WNL/B816. 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

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	—	✓	✓	✓/χ
Regional lesion volume				
Reference e23	—	✓	χ	—
Global GM, longitudinal studies				
Global lesion volume				
Reference e24	—	✓	χ	—
Cortical GM, cross-sectional studies				
Global lesion volume				
Reference e2	χ	✓	—	—
Reference e16	—	✓	✓	—
Reference e17	—	✓	χ	✓/χ
Reference e31	—	✓	—	χ
Reference e34	—	✓	χ	✓
Reference e40	—	χ	✓	✓
Regional lesion volume				
Reference e9	✓	✓	—	—
Reference e46	—	✓	✓	✓
Reference e23	—	χ	χ	—
Reference e45	—	✓	✓	—
Reference e40	—	—	✓	✓
Cortical GM, longitudinal studies				
Global lesion volume				
Reference e10	χ	χ	χ	χ
Deep GM, cross-sectional studies				
Global lesion volume				
Reference e2	✓	✓	—	—
Reference e17	—	✓	✓	✓/χ
Reference e53	—	✓	✓/χ	—
Reference e52	—	χ	✓	—
Reference e16	—	✓	✓	—
Reference e40	—	✓	✓	✓
Reference e55	—	✓	χ	χ

Continued

Table 2 Studies Including Multiple Disease Phenotypes *(continued)*

Study	CIS	RRMS	SPMS	PPMS
Regional lesion volume				
Reference e23	—	χ	χ	—
Reference e45	—	✓	✓	χ
Reference e46	—	✓	✓	✓
Reference e40	—	—	✓	✓
Deep GM, longitudinal studies				
Global lesion volume				
Reference e10	χ	✓	χ	χ

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), χ (association not present), or ✓/χ (association present for some analyses, not present for other analyses). If the disease phenotype was not investigated in the study, this is indicated as —.

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, links.lww.com/WNL/B816.

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.

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

Table 3 Main Findings

No. 1: More WM lesions, more GM atrophy
In cross-sectional studies in particular, 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.
No. 2: WM lesions are most clearly linked to GM atrophy in RRMS
The most consistent relationship 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 relationship with deep GM and especially thalamus volumes was particularly consistent in RRMS and 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.

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.

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

Table 4 Research Aims and Suggestions for Future Research

Research aims	Suggestions
Understand the details of the spatiotemporal 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 interstudy and intrastudy variability in imaging studies and include acquisitions needed to detect and study relevant pathology (e.g., gray matter lesions). Combining or interpreting results of imaging studies in context with knowledge obtained through histopathologic or molecular studies.
Untangle the neurodegenerative processes secondary to focal inflammatory damage from those primarily arising in the gray matter.	
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 pathologic substrate.[30]
Develop therapeutic interventions targeting specific neurodegenerative processes.	Consider the type and duration of therapeutic interventions that included participants have received. Specify the neurodegenerative pathway targeted in clinical trials; interpret results separately for each disease phenotype; and compare them directly.

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.

Study Funding

The authors report no targeted funding.

Disclosure

I.A. Lie has received research grants from the Meltzer Research Fund, Gerda Meyer Nyquist Guldbrandson & Gert Meyer Nyquists Legat and the Independent Order of Odd Fellows. M.M. Weeda reports no disclosures relevant to the manuscript. R.M. Mattiesing is employed on a research grant from Merck. M.A.E. Mol and P.J.W. Pouwels report no disclosures relevant to the manuscript. F. Barkhof has received compensation for steering/safety committee, activities and consulting services from Roche, Biogen, Merck, Combinostics, Janssen and IXICO. He is co-founder and shareholder of Queen Square Analytics LTD. Ø. Torkildsen has received research grants and speaker honoraria from Biogen, Roche, Novartis, Merck, and Sanofi. L. Bø has received unrestricted research grants to his institution and/or scientific advisory board or speaker honoraria from Almirall, Biogen, Genzyme, Merck, Novartis, Roche, and Teva, and has participated in clinical trials organized by Biogen, Merck, Novartis, Roche, and Genzyme. K.M. Myhr has received unrestricted research grants to his institution; has received scientific advisory board and speaker honoraria from Almirall, Biogen, Genzyme, Merck, Novartis, Roche, and Teva; and has participated in clinical trials organized by Biogen, Merck, Novartis, and Roche. H. Vrenken has received research grants from Pfizer, Merck Serono, Novartis. and Teva; speaker honoraria from Novartis; and consulting fees from Merck Serono, all paid directly to his institution. Go to Neurology.org/N for full disclosures.

Publication History

Received by *Neurology* March 5, 2021. Accepted in final form January 3, 2022.

Appendix Authors

Name	Location	Contribution
Ingrid Anne Lie, MD	Department of Radiology and Nuclear Medicine, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, Location VUmc, the Netherlands; Department of Clinical Medicine, University of Bergen, Norway; NeuroSysMed, Department of Neurology, Haukeland University Hospital, Bergen, Norway	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation
Merlin M. Weeda, MSc	Department of Radiology and Nuclear Medicine, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, Location VUmc, the Netherlands	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation

Appendix *(continued)*

Name	Location	Contribution
Rozemarijn M. Mattiesing, MSc	Department of Radiology and Nuclear Medicine, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, Location VUmc, the Netherlands	Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation
Marijke A.E. Mol, PhD	Medical Library, Amsterdam UMC, Vrije Universiteit Amsterdam, the Netherlands	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Petra J.W. Pouwels, PhD	Department of Radiology and Nuclear Medicine, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, Location VUmc, the Netherlands	Drafting/revision of the manuscript for content, including medical writing for content
Frederik Barkhof, MD, PhD	Department of Radiology and Nuclear Medicine, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, Location VUmc, the Netherlands; Institutes of Neurology and Healthcare Engineering, UCL London, UK	Drafting/revision of the manuscript for content, including medical writing for content
Øivind Torkildsen, MD, PhD	Department of Clinical Medicine, University of Bergen, Norway; Neuro-SysMed, Department of Neurology, Haukeland University Hospital, Bergen, Norway	Drafting/revision of the manuscript for content, including medical writing for content
Lars Bø, MD, PhD	Department of Clinical Medicine, University of Bergen, Norway; Norwegian Multiple Sclerosis Competence Centre, Department of Neurology, Haukeland University Hospital, Bergen, Norway	Drafting/revision of the manuscript for content, including medical writing for content
Kjell-Morten Myhr, MD, PhD	Department of Clinical Medicine, University of Bergen, Norway; Neuro-SysMed, Department of Neurology, Haukeland University Hospital, Bergen, Norway	Drafting/revision of the manuscript for content, including medical writing for content
Hugo Vrenken, PhD, IR	Department of Radiology and Nuclear Medicine, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, Location VUmc, the Netherlands	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation

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.
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 e1-e90 are in the supplement, links.lww.com/WNL/B816.

Relationship Between White Matter Lesions and Gray Matter Atrophy in Multiple Sclerosis: A Systematic Review

Ingrid Anne Lie, Merlin M. Weeda, Rozemarijn M. Mattiesing, et al.

Neurology 2022;98;e1562-e1573 Published Online before print February 16, 2022
DOI 10.1212/WNL.0000000000200006

Supplemental data

Contents

eAppendix 1
Search strategies
PubMed Medline search October 11, 2019.
Yield: 1787 titles
[Mesh] = Medical subject headings
[tiab] = words in title or abstract or author keywords

#1	"Multiple Sclerosis"[Mesh] OR multiple sclerosis[tiab] OR ms[tiab]
#2	(("Gray Matter"[Mesh] OR gray[tiab] OR grey[tiab] OR cortical[tiab] OR subcortical[tiab] OR brain[tiab]) AND ("Atrophy"[Mesh] OR atroph*[tiab] OR volum*[tiab])) OR cortical thickness[tiab]
#3	"White Matter"[Mesh] OR "Nerve Fibers, Myelinated"[Mesh] OR white[tiab] OR (t2[tiab] AND lesion*[tiab]) OR focal lesion*[tiab] OR lesion load[tiab] OR lesion burden[tiab]
#4	#1 AND #2 AND #3

Embase.com search October 11, 2019.

Yield: 1963 titles.

/exp = EMtree keyword with explosion
:ab,ti,kw = words in title or abstract or author keywords
NEXT/n = Requests terms that are within n words of each other in the order specified

1	('multiple sclerosis'/exp OR 'multiple sclerosis':ti,ab,kw OR ms:ti,ab,kw)
2	(('gray matter'/exp OR gray:ti,ab,kw OR grey:ti,ab,kw OR cortical:ti,ab,kw OR subcortical:ti,ab,kw OR brain:ti,ab,kw) AND ('atrophy'/de OR 'brain atrophy'/exp OR 'brain cortex atrophy'/exp OR 'adrenal cortex atrophy'/exp OR atroph*:ti,ab,kw OR volum*:ti,ab,kw) OR 'cortical thickness':ti,ab,kw)
3	'white matter'/exp OR 'white matter lesion'/exp OR 'myelinated nerve'/exp OR white:ti,ab,kw OR (t2 NEXT/3 lesion*):ti,ab,kw OR 'focal lesion*':ti,ab,kw OR 'lesion load':ti,ab,kw OR 'lesion burden':ti,ab,kw
4	1 and 2 and 3
5	4 not 'conference abstract'/it

To remove duplicates, records were imported into Endnote X9.2 (Clarivate Analytics, Philadelphia, PA).

eAppendix 2
Expanded results section

CIS

Eight cross-sectional and four longitudinal studies investigated patients diagnosed with clinically isolated syndrome (CIS). In the longitudinal studies, the follow-up period ranged from two to five and a half years.
The association of lesions with global gray matter (GM) measures were reported in five studies, while cortical and deep GM measures were each considered in seven studies. Four studies reported on regional white matter (WM) lesion measures.

Included studies are described in supplemental eTables 1, 2 and 3, available from Dryad.

Global GM in CIS
In two out of three cross-sectional CIS studies, no significant association was found between global GM volume and either T2[e3] or T1[e4] lesion volume (LV). One study did find such a relation: T2 LV and global GM volume was significantly correlated (r=-0.56, $p<0.020$)[e5]. In this study, 17 patients with CIS were included, and the association was not corrected for age and sex, which was done in the two studies with absent associations.
The longitudinal relationship between global WM lesion measures and global GM atrophy was reported in two studies, and both observed somewhat different, but significant associations. In one of the studies, change in GM fraction over three years correlated with the change in LV (T1: r=-0.307, $p=0.0426$, T2: r=-0.4280, $p=0.0032$), but not with baseline lesion measures[e6]. In the other available study, significant associations were found between global GM percentage change and baseline lesion measures ($p\leq0.004$) and total cumulative number of new/enlarging T2 lesions ($p=0.013$), but not with changes in LV during the 48 months follow-up[e7]. In the first study, only three out of 58 included patients received disease modifying therapy (DMT) in the form of interferon-beta, while in the other study, all 210 patients received intramuscular interferon-beta once a week starting from the study baseline.

Cortical GM in CIS
In cross-sectional studies on patients with CIS, lower cortical GM (CGM) volume showed variable associations with global WM lesion measures.
Two studies observed a significant relation with the presence ($t=2.48$, $p=0.020$)[e8] or volume ($r=-0.49$, $p=0.045$)[e5] of T2 lesions, while three studies did not[e2,e3,e8]. In one of these studies lower pericalcarine cortical volumes were observed in patients with optic neuritis, in the presence of whole brain T2 lesions ($t= 2.48$, $p=0.020$) and T2 lesions in the optic radiation (OR) ($t= 2.24$, $p=0.034$). However, pericalcarine volume and thickness did not correlate with whole brain or OR T2 LV[e8]. Another study reporting on regional WM lesion measures observed no significant correlation with whole-brain average cortical thickness, but did observe such associations for regional cortical thickness measures ($p\leq0.0466$) in vertexwise analyses[e9].
Of the two available longitudinal studies, one study found significant associations between cortical volume change and baseline WM lesion measures ($p\leq0.004$) and the total cumulative number of new/enlarging T2 lesions ($p=0.036$) over the 48-month follow-up, however no such associations were observed for changes in LV[e7]. The second study assessed a novel, more unconventional lesion measure; atrophied T2 LV defined as T2-weighted lesional tissue subsequently substituted by CSF. Over a five-and-a-half-year follow-up period, no association with neither baseline cortical volume nor volume change was found[e10].

Deep GM in CIS

In the five available cross-sectional studies in CIS patients, all except one[e3] showed significant associations between global[e2,e4] and regional[e11,e12] WM LV and total[e2] and regional deep GM (DGM) volumes[e2,e4,e11,e12] (*p*-values ranging from <0.0001 to 0.05). By contrast, no associations with DGM volumes were found for global number of T2-lesions or the presence of Gd-enhancing lesions[e2]. Of the regional DGM volumes investigated, the most consistent relationships were found for the thalamus and hippocampus, where in one study, as much as 47% of the variance in thalamic volume was explained by T1 LV in thalamocortical WM ($p<0.001$)[e12]. This pattern was true considering both global[e2,e4] and regional WM LV[e11,e12].

Longitudinally, one of the available studies, in which all 210 patients were treated with first-line DMTs, found that change in thalamic volume over 48 months was related to global baseline lesion measures ($p\leq0.018$) and the total cumulative number of new/enlarging T2 lesions ($p=0.013$), but not to changes in LV[e7]. Similarly, another study considering regional WM LV, found no significant correlation between changes in ipsilateral LV and changes in hippocampal volume over a follow-up of 24 months[e11]. In this study, treatment status for the 36 included patients was not described.

Finally, atrophied T2 LV (described in the section above) was not associated with neither baseline thalamic volume nor volume change[e10].

RRMS

Overall, 37 cross-sectional and 14 longitudinal studies reported associations between WM lesion measures and GM atrophy in relapsing-remitting MS (RRMS) patients. Of these, 17 reported on the relation of WM lesions with global GM, 29 on that with cortical GM, and 25 on that with DGM measures. A total of 11 studies considered regional WM lesion measures. The follow-up period of the available longitudinal studies ranged from one to five and a half years.

Included studies are described in eTables 1, 2 and 3.

Global GM in RRMS

The great majority of available cross-sectional RRMS studies, i.e., eight out of ten studies, observed a significant association between global GM volume and global WM lesion load. Seven studies observed a significant correlation between global GM volumes and T1[e13,e14] and T2[e13-e18] LV, and abnormal WM[e20] (r-values ranging from -0.32 to -0.726, *p*-values ranging from <0.001 to 0.047). Furthermore, T2[e13,e17,e19] LV was as a significant predictor of GM volume in 3 studies (*p*-values ranging from <0.001 to 0.001). In contrast, two studies considering T2 LV[e21,e22], and another two studies considering Gd-enhancing LV[e13,e14], did not observe a significant correlation with global GM volumes. The four studies included a relative low number of patients (between 21 and 37 patients), no other apparent systematical differences were found between these studies, and studies observing significant associations between lesion measures and global GM volumes.

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 out of 26 WM regions, respectively (r-values ranging from -0.49 to -0.50, $p<0.001$)[e23].

Of the seven longitudinal studies available, four did not find an association between global GM atrophy progression and global WM lesion measures. When considering Gd-enhancing lesion measures obtained at baseline, one study found a significant association (standardized $\beta=-0.28$, $p=0.04$) (follow-up time four years)[e24], while three others did not find that global

GM atrophy progression related to either the presence[e25,e26], number[e26] or volume[e14] of Gd-enhancing lesions (follow-up time ranging from one to two years).
Three out of five studies with a follow-up period between one to four 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 to 0.03). Of note, in one of these studies, the association was only present in the patient group treated with fingolimod (r=-0.43, *p*=0.03), and not in the group treated with natalizumab[e28]. Lastly, two studies did not observe significant associations with global GM atrophy progression, neither for increasing T1[e14], T2[e14,e29] nor Gd-enhancing[e14] LV (follow-up time ranging from 12 to 24 months).

CGM in RRMS
A majority of cross-sectional studies (14 out of 19) considering global WM LV found significant associations. Negative correlations between T2 LV and total cortical volume were reported in five studies (r-values ranging from -0.245 to -0.48, *p*-values ranging from <0.0001 to 0.05)[e2,e15,e30-e32]. Furthermore, global cortical thickness was also found to associate with global WM LV (r-values ranging from -0.294 to -0.55, β-values ranging from -0.03 to -0.357, *p*-values ranging from <0.001 to <0.05)[e17,e30,e33-e35].
A total of six 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 (*p*-values ranging from <0.001 to <0.05). A similar pattern of associations was seen for cortical thickness measures, with cortical thinning in temporal, frontal, parietal and cingulate cortex (*p*-values ranging from <0.0001 to <0.05)[e30,e35,e39].
Although the majority of studies describing cortical measures found significant relationships with WM LV, five studies did not, either for cortical volume[e21,e40-e42] or cortical thickness[e39]. The four studies assessing cortical volumes[e21,e40-e42], included a relatively low number of patients (between 26 and 51), compared to the studies observing significant associations with LV.
In two of the eight cross-sectional studies considering regional distribution of WM lesions, the associations with cortical volume were weak or non-significant[e23,e43]. However, the other six publications demonstrated results suggestive of an anatomical or structural relationship between lesion location and regional cortical volume[e44,e45] and thickness[e9,e35,e46,e47].
Three out of five longitudinal studies considering global WM lesion measures and cortical atrophy found significant relationships between them. The only study analysing baseline WM lesion measures found that baseline T2 LV (annual additional volume loss 0.052% per cm^3 of T2 LV, *p*<0.0001) and Gd-enhancing lesion number (annual additional volume loss of 0.046% for each additional Gd-enhancing lesion, *p*=0.0102) were significant predictors of on-study total cortical volume loss[e49]. In the four studies assessing changes in global WM measures, two found an association between increasing lesion volumes[e27,e48] or numbers[e48], and regional volume loss (*p*<0.01)[e27] and cortical thinning (*p*=0.040)[e48] (follow-up time ranging from 1 to 2 years).
These results were in contrast to the third study considering global WM measures, in which the change in T2 LV did not associate with the change in cortical volume over the two-year follow-up[e32]. There were no apparent systematical differences in methodology or clinical characteristics between studies where associations were present and studies where associations were absent. Lastly, one study assessed atrophied T2 LV (described in previous section), and found no association with baseline cortical volume, or volume change[e10].
Two studies assessed the relationship between regional WM LV and cortical atrophy, both by visual inspection. One of them observed that the increase in T2 LV over the follow-up spatially coincided with areas of cortical decrease[e50], while the other study did not[e51].

DGM in RRMS

With the exception of one study[e52], all 17 cross-sectional publications reporting on global WM LV and DGM, found significant associations between the two. Three studies evaluated DGM volume as a whole (β-values ranging from -0.258 to -0.49, *p*-values <0.0001 to 0.04)[e17,e49,e53], while the remaining evaluated the various structures separately. Thalamic volume correlated negatively with T1[e36,e37] and T2[e2,e30,e36-e38,e40,e41] LV in seven studies (r-values ranging from -0.407 to -0.81, *p*-values ranging from <0.05 to <0.00001), and an additional three studies reported significant associations in regression analyses (*p*-values ranging from <0.0001 to <0.05)[e16,e49,e55]. Lastly, one study found that surface displacement of the thalamus and other DGM structures associated with T2 LV (p=0.01)[e54]. Other DGM structures repeatedly showing significant associations with WM LV were the caudate nucleus (*p*-values ranging from <0.0001 to <0.05)[e2,e19,e36-e38,e41,e42,e55,e56], putamen (r-values ranging from -0.176 to -0.57, *p*-values ranging from <0.00001 to <0.05)[e2,e30,e36-e38,e53,e55], and globus pallidus (*p*-values ranging from <0.0001 to <0.05)[e2,e30,e38,e54,e55].

While two cross-sectional studies, one of which only conducted a qualitative evaluation by visual inspection[e44], did not find any associations between regional WM lesion and DGM measures[e23,e45], the majority of studies considering these structures did[e43,e44,e46,e54].

All four publications[e10,e48,e49,e57] that assessed longitudinal relations between total and regional DGM atrophy and global WM lesion measures, observed significant associations (*p*-values ranging from <0.0001 to 0.0372), and most consistently with thalamic atrophy.

Two studies observed that baseline T2 LV associated significantly with thalamic (r=-0.586, p=0.027[e57], β=-0.058, p<0.0001[e49]), and DGM (β=-0.053, p<0.0001)[e49] volume loss over 24 months. One of these studies also found baseline Gd-enhancing lesion number to be a significant predictor of DGM (β=-0.060, p=0.0007) and thalamic (β=-0.039, p=0.0372) atrophy[e49]. In the third available study, DGM atrophy rates were higher in patients with MRI activity (new/enlarging T2 lesions or new Gd-enhancing lesions) during the 24-month follow-up (mean atrophy rate of -0.4.51, p=0.024)[e48]. Finally, in the one study considering atrophied T2 LV (described in previous section), a significant association was found for baseline thalamus volume (r=-0.384), and volume change (r=-0.430, both *p*-values=0.004)[e10].

SPMS

Eleven cross-sectional and two longitudinal studies reported on patients with secondary progressive MS (SPMS), five of which explored associations between WM lesions and global GM volume, while eight and ten studies focused on cortical and DGM measures, respectively. Four studies considered regional WM lesion measures.

Included studies are described in eTables 1 and 2.

Global GM in SPMS

Three out of four cross-sectional studies showed significant negative associations between WM LV and global GM volume (r-values ranging from -0.46 to -0.72, all *p*-values <0.001)[e16,e58]. In one paper, T2 LV did correlate with normalized GM volume (r=-0.36, p<0.01), but did not remain as a significant predictor in the final regression model[e17]. The one 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 fraction over the subsequent four-year follow-up of 19 patients with SPMS[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 available four studies. Two studies found significant associations with lower cortical volume, mainly in frontal, temporal, cingulate and cerebellar regions (p-values <0.001[e40] and <0.05[e16]). Furthermore, cortical thickness was evaluated in another two studies based on one and the same study population: neither study found any significant association between T2 LV and global mean cortical thickness[e17,e34].
In the four studies investigating regional WM lesions, results were also variable. Measured qualitatively by visual inspection, one study found that the distribution of T2 lesions were spatially close to regions with lower GM volume[e45], while another two studies found weak[e46] or non-significant[e23] associations. Lastly, one study observed relatively strong correlations between lower cortical volume in a given lobe, and T2 LV in the same or adjacent lobes (r-values ranging from -0.67 to -0.79, p<0.001)[e40].
Longitudinally, the only available study assessed atrophied T2 LV (described in previous section), and combined patients with SPMS and primary progressive MS (PPMS) into one progressive MS group (PMS) of 42 patients: no relation to either baseline cortical volume or volume changes were found (follow-up time five and a half years)[e10].

DGM in SPMS
In the six 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 four however, T1[e53] or T2[e17] LV associated significantly with both total DGM volume (r=-0.65[e17], β=-0.385[e17] and -1.11[e53], p-values ranging from <0.001 to 0.04), and separate DGM structures like the hippocampus[e52], thalamus and caudate nucleus[e40] (r-values ranging from -0.69 to -0.88, p-values ranging from<0.001 to 0.018). Two out of four studies that included results for regional WM LV or distribution, and the relationship with DGM volume, showed significant associations (r-values ranging from -0.64 to -0.87, p<0.001[e40], average standardized β=-0.264, p<0.05[e46]). Two studies did not find such associations[e23,e45].
There were no apparent systematic methodological or clinical differences in the discrepant studies, neither for those assessing global nor regional WM lesions.
As described in the above section, the only identified longitudinal publication studied atrophied T2 LV in PMS, and found no relation to baseline thalamus volume or volume change[e10].

PPMS
The relationship between WM lesions and various GM measures in patients with PPMS was assessed in 11 cross-sectional studies, and longitudinally in two studies. Three studies performed analyses involving global GM volume, while cortical and DGM measures were each considered nine studies. Three studies considered regional WM LV or distribution.

Included studies are described in eTables 1 and 2.

Global GM in PPMS
The cross-sectional associations between WM lesion load and global GM volume in patients with PPMS were variable. One study reported a significant correlation (r= -0.68, p<0.001) with T2 LV, but in a multiple regression model the association with global GM volume was no longer significant[e17]. In the other available study, no significant associations were found between GM volume and either T2, T1 or gadolinium-enhancing lesion volumes or numbers[e59].
In the available longitudinal study, no association was found between baseline WM lesion measures, and GM volume change over 12 months[e17].

CGM in PPMS
Cross-sectional results for global WM lesion measures and cortical GM 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 to <0.01)[e40] cortical volume and total cortical thickness (standardized β=-0.425, $p<0.05$)[e34]. In the three studies with non-significant results, no associations were found for either cortical volume[e31,e62] or thickness[e17].
Out of three publications assessing regional WM lesion measures, one found a significant association between cortical volume and regional WM lesion measures in anatomically connected areas (r-values ranging from -0.83 to -0.91, $p<0.001$)[e40], while in the other two the associations with cortical thickness or volume were weak[e46] or absent[e45]. Again, there were no apparent systematic methodological or clinical differences in the studies observing significant and non-significant observations, and in all studies the number of included patients were relatively low (between 18 and 31).
Only one longitudinal study was identified (described in previous section), finding no associations between atrophied T2 LV and baseline cortical volume, or volume change[e10].

DGM in PPMS
All but one[e55] of the six cross-sectional studies reporting on the relationship between global WM lesions and DGM volume observed significant associations. In PPMS patients, correlations were significant for both DGM volume as a whole (r-values ranging from -0.651 to -0.71, p-values ranging from <0.001 to <0.01)[e17,e61], and for 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 to <0.05)[e40,e61,e63] and T1 LV (r-values ranging from -0.44 to -0.554, p-values ranging from 0.002 to <0.05)[e61,e63]. Of note, in the only study with absent associations[e55], 25 patients with both SPMS and PPMS were pooled in a combined PMS group.
Three cross-sectional publications analysed regional WM lesions and the relationship with DGM volume. Lower thalamus volume were related to T2 LV in all lobes analysed in one study (r-values ranging from -0.85 to -0.93, $p<0.001$)[e40], and another study found that regional T2 LV was a significant factor in multiple regression models for the nucleus accumbens, hippocampus and globus pallidus, but not for amygdala, putamen or thalamus (average standardized β=-0.438, $p<0.05$)[e46]. The third study conducted a qualitative analysis by visual inspection, and found no spatial correspondence between T2 lesions and lower DGM volume[e45].
Lastly, only one longitudinal study was available (described in previous section), and for both baseline thalamic volume and volume change, no relation to atrophied T2 LV was found[e10].

Results for mixed MS groups
In a mixed MS patient group, i.e., comprising different disease types but studying the entire patient group as a whole, 32 publications assessed the cross-sectional, and 9 the longitudinal association between WM lesion measures and GM volume. In this section, studies including patients with CIS in the total patient population are also considered. The follow-up time in the longitudinal studies ranged from two to five and a half years.
Global GM volume and its relation to WM lesions was described in 13 studies, CGM volume or thickness was considered in 17, and DGM volume in 23 studies. Associations with regional WM lesions were reported in eight studies in total.

Included studies are described in eTables 1, 2 and 3.

Global GM in mixed MS groups

In all but one of the nine available cross-sectional studies, global WM lesion measures were consistently associated with global GM volume in patients with MS. Significant correlations were found in seven studies (r-values ranging from -0.43 to -0.63, *p*-values ranging from <0.001 to 0.016)[e16,e17,e65-e69], and two studies observed WM LV as a significant predictor of global GM volume (r=-0.52[e17], β=-0.226[e17] and -0.27[e1], *p*-values ranging from <0.01 to <0.001). In all of the above studies, patients with RRMS made up the majority of the mixed MS groups. Patients with or without Gd-enhancing lesions however, did not exhibit different GM volumes[e64].

One cross-sectional study reported results on regional WM LV, and observed significant correlations between total GM fraction and regional T2 and T1 LV in 9 and 5 out of 26 regions, respectively (r-values ranging from -0.24 to -0.45, $p<0.001$)[e23].

The four available longitudinal studies considering global GM atrophy and global WM lesion measures load reported mainly absent associations. In the three studies assessing baseline WM lesion measures, changes in GM volume were not associated with the presence of Gd-enhancing lesions[e64], T1 LV[e88] or abnormal WM fraction[e89] (follow-up period ranging from two to five and a half years). However, in one of these studies, baseline global GM volume correlated significantly with variation in abnormal WM fraction over the 2-year follow-up (r=-0.180, $p<0.001$)[e89]. Lastly, in the one study assessing T2 lesion shrinking over a three-year follow-up, this did not relate to changes in global GM volume[e87].

CGM in mixed MS groups

The majority of studies, i.e., ten out of twelve studies considering cortical volume or thickness and global WM lesion measures in MS, found significant associations between the two. T2 LV was associated with lower GM volume in the cerebellar cortex, temporal lobe and the pre- and postcentral gyrus in the frontal and parietal lobe in two studies (both *p*-values <0.05)[e16,e73]. One study using estimated regional cortical, DGM and brainstem volumes in an event-based model to determine the sequential occurrence of atrophy, observed that the event-based model stage at baseline was related to the T2 LV ($p<0.001$)[e71]. Five studies found a significant association between global WM LV and both global (*p*-values ranging from <0.001 to <0.05)[e17,e34,e70,e75] and regional cortical thickness (*p*-values ranging from 0.007 to <0.05)[e74,e75,e77]. Lastly, one study found that global T2 LV correlated significantly with cortical surface area (r= -0.62, $p<0.05$)[e72].

In two publications, one of which all included patients were treated with natalizumab, cortical thickness and/ or volume did not associate with either global T2 LV[e76] or the presence of Gd-enhancing lesions[e64].

Associations between regional WM lesion measures and cortical atrophy were reported in five available studies. One study found that regional T2 LV in the left and right frontal, temporal, parietal, and occipital areas were associated with lower cortical thickness in widespread bilateral cortical regions ($p<0.05$)[e77], and another study observed that LV in connected WM tracts was a significant explanatory variable of cortical thickness (average standardized β=-0.116, $p<0.05$) in 28 out of 34 areas[e46]. In the three remaining publications, associations between regional WM lesion measures and cortical volume or thickness were weak (r=-0.267, $p\leq0.01$)[e23] or non-significant[e9,e78].

Four publications explored the longitudinal relationship between global WM lesion measures and cortical atrophy, and none of them observed significant associations. No relations were found between the presence of Gd-enhancing lesions at baseline[e64] or changes in T2 LV[e77], and changes in cortical volume[e64] or thickness (either global or regional)[e64,e77] (follow-up time ranging from 3 to 4 years). In one study using an event-based model (described in the above

section); the rate of increase in T2 LV over the mean follow-up of 2.4 years did not associate with the rate of change in the event-based model stage[e71]. Assessing atrophied T2 LV (described in previous section), associations with baseline cortical volume or volume change were not found[e10].
One of the four longitudinal publications also considered regional T2 LV, and found that changes in left temporal and occipital T2 LV associated negatively with cortical thinning in the left temporal, parietal and occipital areas ($p<0.05$)[e77].

DGM in mixed MS groups
Seventeen studies considered the cross-sectional relationship between global WM lesion and DGM measures in patients with MS; as many as 14 observed significant associations. Global T1[e53] and T2[e17] LV was significantly associated with total DGM volume (r=-0.70[e17], standardized β= -0.407[e17] and -0.41[e53], *p*-values ranging from <0.001 to 0.02) as well as with the separate structures. A total of ten publications reported significant associations between global WM LV and thalamic (r-values ranging from -0.36 to -0.77, *p*-values ranging from <0.0001 to <0.05)[e16,e53,e55,e76,e80,e81,e83-e85] and pulvinar nucleus ($p<0.05$) volume[e73].
Associations were also seen for caudate nucleus (*p*-values ranging from <0.01 to <0.05)[e16,e55,e73,e76], putamen (*p*-values ranging from 0.003 to <0.05)[e53,e55,e69,e73] and hippocampal volume (r-values ranging from -0.46 to -0.56, all *p*-values=0.008)[e82].
Furthermore, global T2 LV was associated with a more advanced event-based model stage, as referred to in the above section regarding cortical GM atrophy in MS patients ($p<0.001$)[e71].
As opposed to the results listed above, two studies did not observe significant associations between WM LV and lower GM volume either in the caudate nucleus[e79] or in any other DGM structure[e74]. Furthermore, the presence of Gd-enhancing lesions did not correlate with total or regional DGM volume[e64].
Of the available studies analysing the cross-sectional relationship between regional WM LV and DGM volume, three found significant (*p*-values ranging from 0.02 to <0.05)[e46,e73,e78], or near significant associations[e23,e86].
Only one out of three studies[e10,e64,e71] reporting on the longitudinal relationship between DGM atrophy and global WM lesion measures in MS patients found a significant association: in this study, atrophied T2 LV correlated significantly with baseline thalamic volume (r=-0.620, p=0.003) and thalamic volume change (r=-0.672, p=0.003)[e10]. The presence of Gd-enhancing lesions at baseline did not associate with changes in total or regional DGM volume over a 3-year follow up, in patients treated with natalizumab[e64]. Furthermore, the rate of change in the event-based model stage was not related to the rate of increase in T2 LV[e71].
Two longitudinal studies considering regional WM lesion measures and DGM atrophy were available, and both found significant associations. One study found that over a 12-month period, new, chronic enlarging and chronic shrinking T1 lesion number along the optic radiation was related to reduced ipsilateral lateral geniculate nucleus volume (*p*-values ranging from 0.0001 to 0.0056)[e90]. The second study assessed regional DGM atrophy over five years, and its association to percentage lesion disruption, i.e., the percentage of normative data-base derived connected tract streamlines that passes through a given lesion mask and are considered disrupted. When controlling for T2 LV, tract disruption could predict DGM volume atrophy rate in 5 of 14 regions, but only 1 after correction for multiple comparisons ($p<0.001$)[e86].

Comparisons between disease phenotypes
Four studies described global GM atrophy, while eleven studies reported results on CGM and DGM atrophy, each. Five studies considered regional WM lesion measures and their possible association with GM atrophy.

The included studies are described in eTables 1 and 2.

Global GM in comparisons between phenotypes
In both available cross-sectional studies, a negative association between global GM volume and global WM LV was observed in all investigated disease types (RRMS (r=-0.334, $p<0.001$[e16], r=-0.57, $p<0.01$[e17]), SPMS (r=-0.460, $p<0.001$[e16], r=-0.36, $p<0.01$[e17]), PPMS (r=-0.68, $p<0.01$[e17])), although one of these found that in multivariable logistic regression, the association with global WM lesion volume remained only for RRMS (standardized β=-0.231, $p<0.01$), but not SPMS or PPMS[e17].
In the single study assessing regional WM LV, significant associations were only found in RRMS (r-values ranging from -0.20 to -0.50, $p\leq0.001$), but not in SPMS[e23].
Longitudinally, the presence of Gd-enhancing lesions at baseline (standardized β=-0.28, $p=0.04$) and increase in T2 LV during the four-year follow-up (standardized β=-0.46, $p=0.0004$) predicted faster GM fraction change in RRMS, but not in SPMS[e24].
In all of the above studies, the RRMS group was considerably larger than the groups consisting of patients with SPMS or PPMS.

CGM in comparisons between phenotypes
The six available cross-sectional studies found more consistent associations between global WM lesion load and cortical atrophy in RRMS than the other phenotypes, though the association was not uniformly present in all phenotypes[e2,e16,e17,e31,e34,e40]. Four studies found that global T2 LV was associated with total cortical volume or thickness in RRMS (r-values ranging from -0.420 to -0.55[e2,e17,e31], standardized β=-0.319[e34] and -0.357[e17], *p*-values ranging from <0.0001 to <0.05), while in CIS it was absent[e2], in SPMS absent[e17,e34], and in PPMS present (r=-0.43, $p<0.05$[e17], standardized β=-0.425, $p<0.05$[e34]) or absent[e17, e31] (in one of the studies[e17] depending on the statistical analysis used); of note, two of these studies investigated the same patient group[e17, e34]. Except the one study including patients with CIS and RRMS[e2], the patient groups were again unbalanced with large RRMS groups compared to groups with progressive disease types. Regional cortical volumes associated with global T2 LV in both RRMS and SPMS in one study (all *p*-values <0.05)[e16], but only in SPMS (r-values ranging from -0.63 to -0.75), and PPMS (r-values ranging from -0.83 to -0.85), not RRMS, in another study (all *p*-values<0.001)[e40].
In the five cross-sectional studies considering regional WM LV, two assessed cortical thickness, finding the most consistent and widespread associations in RRMS patients (*p*-values ranging from 0.0002 to <0.05)[e9,e46]. In CIS[e9], SPMS[e46] and PPMS[e46], the association was significant only in certain cortical regions. In the three studies assessing cortical volume, associations were absent in RRMS and SPMS[e23], present in RRMS and SPMS (qualitative association by visual inspection)[e45] and present in SPMS (r-values ranging from -0.65 to -0.79, $p<0.001$) and PPMS (r-values ranging from -0.83 to -0.91, $p<0.001$)[e40].

DGM in comparisons between phenotypes
All three cross-sectional studies assessing the relationship between global WM lesion load and total DGM volume, observed a significant association in RRMS (r=-0.613[e2] and -0.74[e17], standardized β=-0.407[e17] and -0.49[e53], *p*-values ranging from <0.0001 to 0.04) while finding it to be present (r=-0.329, $p<0.0001$)[e2] in CIS, absent[e53] or present (r=-0.65[e17], standardized β=-0.385[e17] and -1.11[e53], *p*-values ranging from <0.001 to 0.04) in SPMS (in one of the studies[e53] depending on the lesion measure), and absent[e17] or present (r=-0.71, $p<0.001$)[e17] in PPMS (depending on the statistical analysis used[e17]). The two papers considering patients with RRMS and progressive disease types[e17,e53] had unbalanced patient groups with small groups

of SPMS (53[e17] and 12[e53] patients) and PPMS (25 patients[e17]) patients. In the six studies considering regional DGM volumes, global WM lesion measures were, with the exception of one study[e52], consistently related in RRMS (*p*-values ranging from <0.0001 to 0.04)[e2,e16,e40,e53,e55]. Furthermore, associations were present in CIS (r-values ranging from -0.218 to -0.328, *p*-values ranging from <0.0001 to <0.001)[e2], absent[e16,e55] or present (r-values ranging from -0.69 to -0.88, *p*-values ranging from <0.001 to 0.018)[e40,e52] in SPMS, and absent[e55] or present (r=-0.94, *p*<0.001)[e40] in PPMS.

In the four cross-sectional studies assessing the relationship between regional WM lesion measures and regional DGM volume, significant associations were absent in RRMS and SPMS[e23], present in RRMS and SPMS, but absent in PPMS (qualitative association by visual inspection)[e45], present in RRMS, SPMS and PPMS (standardized β-values ranging from -0.264 to -0.438, all *p*-values <0.05)[e46], and present in SPMS and PPMS (r-values ranging from -0.64 to -0.87, all *p*-values<0.001)[e40].

Supplementary eReferences

e1. 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.

e2. 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. *AJNR Am J Neuroradiol.* 2012;**33**(8): p. 1573-1578.

e3. 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.

e4. 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.

e5. Labiano-Fontcuberta A, Mato-Abad V, Alvarez-Linera J et al. Gray Matter Involvement in Radiologically Isolated Syndrome. *Medicine*. 2016;**95**(13): p. e3208.

e6. 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.

e7. 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.

e8. 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.

e9. 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.

e10. 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.

e11. 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.

e12. 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.

e13. 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.

e14. 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.

e15. 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.

e16. 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.

e17. 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.

e18. 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.

e19. 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.

e20. 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.

e21. 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.

e22. 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.

e23. 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.

e24. Fisher E, Lee J.C, Nakamura K, Rudick R.A. Gray matter atrophy in multiple sclerosis: A longitudinal study. *Ann Neurol.* 2008;**64**(3): p. 255-265.

e25. 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.

e26. 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.

e27. 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.

e28. 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;**17**(1):p.208-217.

e29. 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.

e30. 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.

e31. 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.

e32. 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.

e33. 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.

e34. 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.

e35. 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.

e36. 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.

e37. 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.

e38. 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.

e39. 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.

e40. 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.

e41. 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.

e42. 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.

e43. 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.

e44. 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.

e45. 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.

e46. 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.

e47. 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.

e48. 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.

e49. 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.

e50. 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.

e51. 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.

e52. 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.

e53. Kalinin I, Makshakov G, Evdoshenko E. The Impact of Intracortical Lesions on Volumes of Subcortical Structures in Multiple Sclerosis. *AJNR Am J Neuroradiol*. 2020. doi: https://doi.org/10.3174/ajnr.A6513

e54. 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.

e55. 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.

e56. 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.

e57. 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.

e58. 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.

e59. 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.

e60. 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.

e61. 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.

e62. 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.

e63. 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.

e64. 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.

e65. 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.

e66. 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.

e67. 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.

e68. 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.

e69. 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.

e70. 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.

e71. 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.

e72. Hier D.B, Wang J. Reduced cortical surface area in multiple sclerosis. *Neurol Res*. 2007;**29**(3): p. 231-2.

e73. 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.

e74. 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.

e75. 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.

e76. 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.

e77. Tsagkas C, Chakravarty M.M, Gaetano L et al. Longitudinal patterns of cortical thinning in multiple sclerosis. *Hum Brain Mapp*. 2020;**41**(8): p. 2198-2215.

e78. 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.

e79. 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.

e80. 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.

e81. 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.

e82. 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.

e83. 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.

e84. 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*. 2021 Apr;27(5):674-683

e85. 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.

e86. 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.

e87. 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.
e88. 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.
e89. 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.
e90. 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.

Supplemental eTable 1. Characteristics of cross-sectional studies

Cross-sectional studies-sectional studies					
Reference (e-ref)	*Patients (n)*	*GM atrophy measure*	*WM lesion measure*	*Results*	*Quality of evidence*
CIS					
Durhan et al., Int J Neurosci, 2016 (e3)	33	Total GM, global volume. Cortical GM, global volume. Cortical GM, regional volume. Deep GM, regional volume.	Global T2 lesion volume.	Significantly higher volume in the left cerebellar cortex, left accumbens area and right GP in smoking CIS patients than in non-smoking CIS patients, but no significant correlation between T2 LV and GMV in these areas.	Fair
Henry et al., J Neurol Neurosurg Psychiatry, 2008 (e4)	41	Total GM, global volume. Deep GM, regional volume.	Global T1 lesion volume.	T1 LV was negatively associated with bilateral thalami (R^2=0.50, p=0.020 (left) and p=0.037 (right)) and left hippocampus (R^2=0.38, p=0.037) volume[a]. No significant association between T1 LV and global GMV.	Good
Henry et al., J Neurol Sci, 2009 (e12)	24	Deep GM, regional volume (thalamus).	Regional T1 lesion volume.	In a stepwise regression model, 47% of the variance in thalamic volume was explained by T1 LV in thalamocortical WM (p<0.001)[a].	Good
Jenkins et al., J Neurol Neurosurg Psychiatry, 2011 (e8)	28	Cortical GM, regional volume and thickness (pericalcarine cortex).	Global T2 lesion volume. Regional T2 lesion volume (optic radiation). Global presence of T2 lesions Regional presence of T2 lesions (optic radiation). Regional T2 lesion length (optic nerve).	Lower pericalcarine cortical volumes were significantly associated with the presence of whole brain T2 lesions (t=2.48, p=0.020) and OR T2 lesions (t=2.24, p=0.034), in patients with optic neuritis[a]. No significant association between pericalcarine volume or thickness and optic nerve lesion length, OR T2 LV or whole brain T2 LV.	Fair
Labiano-Fontcuberta et al., Medicine, 2016 (e5)	17	Total GM, global volume. Cortical GM, global volume.	Global T2 lesion volume.	Significant negative correlation between T2 LV and GMV (r=-0.56, p<0.020), and CV (r=-0.49, p<0.045)[a].	Fair
RRMS					
Al-Radaideh et al., Clin	30	Cortical GM, global volume, thickness and surface area.	Global T2 lesion volume.	Significant correlation between T2 LV and cortical surface area (r=-0.208, p<0.05), Cth (r=-0.294,	Fair

Neuroradiol, 2019 (e30)		Cortical GM, regional volume, thickness and surface area. Deep GM, regional volume.		$p<0.05$), and cortical ($r=-0.245$, $p<0.05$) putamen ($r=-0.408$, $p<0.05$), GP ($r=-0.410$, $p<0.05$) and thalamus ($r=-0.407$, $p<0.05$) volumes[a].	
Bergsland et al., Mult Scler, 2015 (e47)	51	Cortical GM, regional thickness and surface area (primary motor cortex).	Regional T2 lesion volume (corticospinal tract).	Significant correlation between CST T2 LV and PMC thickness ($r=-0.327$, $p=0.022$)[a], but not with PMC surface area ($r=-0.039$, $p=0.791$). Significant correlation between PMC thickness and lesion probability along large parts of the CST (p-values ranging from 0.004 to 0.048)[a]. CST T2 LV not a significant predictor of PMC thickness in final multiple regression model.	Fair
Calabrese et al., Neurology, 2010 (e33)	100	Cortical GM, global thickness.	Global T2 lesion volume.	Significant correlation between T2 LV and mean Cth ($r=-0.31$, $p=0.004$)[a].	Good
Ceccarelli et al., J Neurol, 2007 (e21)	32	Total GM, global volume. Cortical GM, global volume.	Global T2 lesion volume.	No significant correlation between total T2 LV and any GM volume measure.	Fair
Chard et al., Brain, 2002 (e13)	26	Total GM, global fraction (to ICV).	Global T1, T2 and Gd-enhancing lesion volume.	Significant correlation between GMF and T1 LV ($r=-0.525$, $p=0.006$) and T2 LV ($r=-0.726$, $p<0.001$), but not with Gd-enhancing LV ($r=0.109$, $p=0.595$)[a]. In a general linear model, GMF decreased by -0.003 for every ml T2 LV (SE=0.001, $p<0.001$).	Fair
Charil et al., Neuroimage, 2007 (e35)	425	Cortical GM, global thickness. Cortical GM, regional thickness.	Global WM lesion volume. Regional WM lesion volume.	Significant association between global WM LV and mean Cth ($t=-15.3$, $p<0.0001$), with an average loss of 0.03 mm in Cth per cm^3 of WM LV, accounting for 38% of the variance[a]. Regionally, global WM LV was associated with Cth in almost all vertices in the surface model, e.g., bilateral cingulate gyrus, prefrontal, temporal and parietal association cortical regions (average loss 0.03 to 0.07 mm per cm^3, t-values ranging from -11.7 to -16.2, all $p<0.05$)[a]. An average lesion density map showed that anterior cortical thinning correlated more with anterior lesions than posterior cortical thinning correlated with posterior lesions[a].	Fair

Datta et al., Mult Scler Relat Disord, 2015 (e36)	924	Cortical GM, regional volume. Deep GM, regional volume.	Global T1, T2 and Gd-enhancing lesion volume.	Significant negative correlation between T2 and T1 LV and thalamus (T2:r=-0.492, T1:r=-0.473), putamen (T2:r=-0.193, T1:r=-0.176), CN (T2:r=-0.188, T1:r=-0.173), inferior frontal gyrus (T2:r=-0.120, T1:r=-0.101), septal nuclei (T2:r=-0.146, T1:r=-0.153), and amygdala (T2:r=-0.109) volumes[a]. T2 and T1 LV correlated positively with hippocampus volume (T2:r=+0.117, T1:r=+0.169)[a]. All p<0.05.	Fair
Dolezal et al., Int Rev Neurobiol, 2007 (e15)	41	Total GM, global volume. Cortical GM, global volume.	Global T2 lesion volume.	Significant correlation between global T2 LV and GMV (r=0.4, p=0.003) and CV (r=0.48, p=0.001)[a]. Higher total T2 LV was associated with lower CV (R^2=0.17, p=0.007)[a].	Fair
Duan et al., Eur J Radiol, 2012 (e41)	26	Cortical GM, regional volume. Deep GM, regional volume.	Global T2 lesion volume.	Significant correlation between T2 LV and the right CN (r=−0.409, p=0.019), left thalamus (r=−0.596, p=0.001) and right thalamus (r=−0.694, p=0.000) volume[a].	Fair
Hasan et al., J Magn Reson Imaging, 2009 (e56)	32	Deep GM, regional volume percentage (of ICV) (caudate nucleus).	Global T2 lesion volume percentage (of ICV).	Significant correlation between global T2 LV percentage and CN volume percentage (r=-0.482, p=0.005)[a].	Fair
Hasan et al., J Neurol Sci, 2012 (e38)	54	Cortical GM, regional volume percentage (of ICV). Deep GM, regional volume percentage (of ICV).	Global T2 lesion volume.	Significant age-adjusted correlation between global T2 LV and corpus striatum (r=-0.519, p<0.0001), thalamus (r=-0.569, p<0.00001), CN (r=-0.46, p=0.001), putamen (r=-0.494, p<0.00001), GP (r=-0.395, p=0.003), accumbens area (r=-0.407, p=0.002) and temporal cortex (r=-0.315, p=0.022) volume percentage[a].	Fair
Kuceyeski et al., AJNR, 2015 (e43)	121	Cortical GM, regional volume. Deep GM, regional volume.	Change in Connectivity (ChaCo) score (the percentage of tracts connecting to a GM region that pass through the WM abnormality mask).	Out of 86 GM regions, no significant correlation was found between regional ChaCo score and lower GMV. Significant correlation between ChaCo score and bilateral thalami (r=0.30), putamen (r=0.38), CN (r=0.26), GP (r=0.26) and nucleus accumbens (r=0.28) volume[a]. No significant correlation between regional lower GMV and ChaCo score in temporal poles.	Fair

Magon et al., HBM, 2014 (e54)	118	Deep GM, regional shape.	Global T2 lesion volume. Regional T2 lesion volume.	Global T2 LV significantly related to surface displacement of the thalamus (F(3, 114)=2.31, *t*=3.12), striatum (F(3, 114)=1.14, *t*=2.79) and GP (F(3, 114)=0.16, *t*=3.21)[a]. Observed displacements in thalamus (F(3,114)=5.44, *t*=3.59), GP (F(3,114)=5.16, *t*=3.41) and striatum (F(3,114)=5.44, *t*=3.59) were mainly driven by frontal and parietal T2 LV (p=0.01)[a], while no surface displacements were associated with occipital T2 LV.	Fair
Narayana et al., Neuroimage Clin, 2012 (e39)	250	Cortical GM, global thickness. Cortical GM, regional thickness.	Global T1 and T2 lesion volume.	No significant correlations between mean global Cth and T1 and T2 LV. Weak to moderate correlations between T1 and T2 LV and regional cortical areas in both 1.5 and 3T, with the strongest correlations seen for the bilateral superior frontal lobe cortex and T1 (*r*=-0.35 (right), *r*=-0,33 (left), p<0.0001) and T2 (*r*=-0.3 (right), *r*=-0.29 (left), p<0.0001) LV[a].	Fair
Prinster et al., Mult Scler, 2010 (e19)	128	Total GM, global volume. Cortical GM, regional volume. Deep GM, regional volume.	Global T2 lesion volume.	Significant linear association between T2 LV and lower global GMV (p=0.001)[a]. T2 LV significantly negatively associated with GMV in the caudate heads (*t*=7.20), parahippocampal gyri (*t*= 5.95), cingulate gyri (*t*=5.86), motor cortex (*t*=5.0), and insula (all p<0.05)[a].	Fair
Prinster et al., Neuroimage, 2006 (e42)	51	Cortical GM, regional volume. Deep GM, regional volume.	Global abnormal WM (aWM).	Significant association between aWM and lower right CN volume (p<0.05)[a]. No significant association between aWM and lower regional CV.	Fair
Quarantelli et al., Neuroimage, 2003 (e20)	50	Total GM, global fraction (to ICV).	Fractional abnormal WM (aWM/ICV = aWMf).	Lower global GMF associated significantly with aWMf (β=-0.62, R=-0.434, p<0.001)[a], with corresponding increase in CSF fraction.	Fair
Riccitelli et al., Mult Scler, 2012 (e44)	78	Cortical GM, regional volume. Deep GM, regional volume.	Regional T2 lesion distribution.	Significant correlation between T2 lesion distribution and lower GMV in the head of the right (*r*=-0.58) and left (*r*=-0.53) CN, right (*r*=-0.39) and left (*r*=-0.43) insula, left cingulate gyrus (*r*=-0.48), right superior	Fair

				frontal gyrus (r=-0.47) and left middle occipital gyrus (r=-0.54)[a].	
Sbardella et al., PLoS One, 2013 (e22)	36	Total GM, global fraction (to ICV).	Global T2 lesion volume.	No significant correlation between T2 LV and GMF.	Fair
Tao et al., J Neurol Sci, 2009 (e37)	88	Cortical GM, regional volume. Deep GM, regional volume.	Global T1 and T2 lesion volume.	Significant correlation between T2 LV and lower volume of the thalamus (r=-0.56), CN (r=-0.31), putamen (r=-0.5), amygdala (r=-0.45), cingulate gyrus (r=-0.47), middle temporal gyrus (r=-0.33), inferior parietal lobule (r=-0.23), entorhinal area (r=-0.34) and insular cortex (r=-0.25)[a]. Significant correlation between T1 LV and lower volume of the thalamus (r=-0.61), CN (r=-0.35), putamen (r=-0.43), pons (r=-0.21), amygdala (r=-0.28), cingulate gyrus (r=-0.43), middle temporal gyrus (r=-0.34), inferior parietal lobule (r=-0.30) and entorhinal area (r=-0.23)[a].	Fair
Toth et al., Front Neuroanat, 2017 (e18)	52	Total GM, global volume.	Global T2 lesion volume.	Significant negative association between T2 LV and GMV (R=-0.32, p<0.021)[a].	Fair
SPMS					
Furby et al., Mult Scler, 2009 (e58)	117	Total GM, global fraction (to ICV).	Global T1 and T2 lesion volume.	Significant negative correlations between GMF and T2 (r=-0.70, p<0.001) and T1 (r=-0.72, p<0.001) LV[a]. Significant associations between GMF and T2 (r_p=-0.69, p<0.001) and T1 (r_p=-0.72, p<0.001) LV in a multiple regression model[a]. In a stepwise multiple regression analysis, T1 LV was the only significant lesion correlate of GMF (standardized β: -0.72, r_p=0.72, R^2=0.52, p<0.001)[a].	Fair
PPMS					
Galego et al., Neuroradiol J, 2015 (e61)	19	Cortical GM, global volume. Cortical GM, regional volume. Deep GM, global volume. Deep GM, regional volume.	Global T1 and T2 lesion volume.	Significant negative correlations between T1 LV and CV (r=-0.508, p<0.05), DGMV (r=-0.651, p<0.01), thalamus (r=-0.554, p<0.05), putamen (r=-0.699,	Poor

				$p<0.01$), GP (r=-0.495, $p<0.05$) and pre-central gyrus (r=-0.605, $p<0.01$) volume[a]. Significant negative correlations between T2 LV and DGMV (r=-0.630, $p<0.01$), thalamus (r=-0.567, $p<0.05$) and putamen (r=-0.696, $p<0.01$) volume[a].	
Mesaros et al., AJNR, 2011 (e63)	54	Deep GM, regional volume (thalamus).	Global T1 and T2 lesion volume.	At baseline, significant negative correlations were found between thalamus volume and T2 LV (r=-0.48, p=0.001) and T1 LV (r=-0.44, p=0.002)[a].	Fair
Sastre-Garriga et al., Neuroimage, 2004 (e59)	43	Total GM, global fraction (to ICV).	Global T1, T2 and Gd-enhancing lesion volume. Global Gd-enhancing lesion number.	No significant correlation between GMF and T2 LV, Gd-enhancing LV or Gd-enhancing lesion number. No significant association between GMF and T2 or T1 LV in regression analyses.	Fair
Sepulcre et al., Arch Neurol, 2006 (e62)	31	Cortical GM, regional volume. Deep GM, regional volume.	Global T2 lesion volume.	Significant associations were found between higher T2 LV and lower local GMV in the anterior nuclei of both thalami (number of voxels, 893; z=4.95, p=0.003)[a], but not with any other cortical or DGM regions.	Fair
MS					
Bermel et al., Neuroreport, 2003 (e79)	24 RRMS (16) SPMS (8)	Deep GM, regional volume (caudate nucleus).	Global T1 and T2 lesion volume.	No significant correlation between CN volume and T1 (p=0.32) or T2 (p=0.23) LV.	Fair
Calabrese et al., J Neurol, 2007 (e70)	83 CIS (10) RRMS (42) SPMS (31)	Cortical GM, global thickness.	Global T2 lesion volume.	Significant correlation between Cth and T2 LV (r=-0.393, p=0.03)[a].	Fair
Deppe et al., Hum Brain Mapp, 2016 (e80)	122 CIS (12) RRMS (110)	Deep GM, regional volume percentage (of ICV) (thalamus).	Global T2 lesion volume.	There was a logarithmic dependency between RTV and T2 LV in the patient groups combined (RTV[%]=1.0028-0.1693 $\log_{10}$T2LV[mL][a], i.e., the higher the T2 LV, the lower the RTV, but the relative impact on thalamic volume is smaller with increasing LV. In a univariate general linear model, T2 LV had the second highest effect (after thalamic FA) on RTV (SS=0.0560, ms=0.0560. F=12.95, $p<0.001$)[a].	Fair

Fisniku et al., Ann Neurol, 2008 (e65)	73 CIS (29) RRMS (33) SPMS (11)	Total GM, global volume. Total GM, global GM fraction (to ICV).	Global T2 lesion volume.	Significant correlation between GMF and T2 LV for the whole cohort (including patients with CIS) of patients (r=-0.63, p<0.001), and for the separate MS (i.e., RRMS and SPMS) subgroup (r=-0.66, p<0.001)[a]. Significant correlation between GMV and T2 LV for the whole cohort (r=-0.57, p<0.001) and for the separate MS (i.e., RRMS and SPMS) subgroup (r=-0.67, p<0.001)[a].	Fair
Fragoso et al., J Clin Neuroscience, 2017 (e66)	185 CIS (1) RRMS (173) PPMS (11)	Total GM, global volume.	Global T2 lesion volume.	Significant negative correlation between T2 LV and GMV (p<0.001)[a].	Fair
Hasan et al., J Neurosci, 2011 (e81)	109 CIS (9) RRMS (88) SPMS (12)	Deep GM, regional volume percentage (of ICV) (thalamus).	Global T2 lesion volume.	Significant correlation between T2 LV and percentage of total thalamus volume (r=-0.658, p<0.001)[a].	Fair
Hier et al., Neurological Research, 2007 (e72)	15 RRMS (6) SPMS (9)	Cortical GM, global surface area.	Global T2 lesion volume.	Negative correlation between T2 LV and cortical surface area (r=-0.62, p<0.05)[a].	Poor
Longoni et al., Brain Struct Funct, 2015 (e82)	103 RRMS (22) SPMS (33) PPMS (23) Benign MS (25)	Deep GM, regional volume (hippocampus).	Global T1 and T2 lesion volume.	Significant correlation between T1 LV and right (r=-0.53, p=0.008) and left (r=-0.46, p=0.008) hippocampus volume[a]. Significant correlation between T2 LV and right (r=-0.56, p=0.008) and left (r=-0.51, p=0.008) hippocampus volume[a]. Significant negative correlation between T1 and T2 LV and local volume of the lateral CA1 subfield, part of the subiculum, and the CA1 region of the hippocampal head, bilaterally (r-values ranging from -0.2 to -0.5; p<0.001; FDR rate at p<0.05 between 4 and 7%)[a].	Fair
Louapre et al., Mult Scler, 2018 (e83)	41 Early RRMS (10) RRMS (18) SPMS (13)	Deep GM, regional volume (thalamus).	Global T2* lesion volume.	Significant partial correlation between thalamic volume and T2* LV (r_p=-0.77, R^2_p=0.59 p=4×10^{-9})[a]. In a multiple regression model, lower thalamic volume associated significantly with higher T2* LV,	Fair

				independently of Cth, cortical LV, ICV, age, sex and disease duration. T2* LV was one of the explanatory variables retained in the model (adjusted R^2=0.68, p=0.8×10^{-5})[a].	
Mehndiratta et al., Mult Scler, 2020 (e84)	90 RRMS (61) SPMS (29)	Deep GM, regional volume (thalamus).	Global T2* lesion volume.	Thalamic volume inversely correlated with T2* LV (r=-0.6, p<0.001)[a]. In multiple regression analysis, lower thalamic volume associated with T2* LV (β=-2.8×10^{-4}, 95% CI:-4.6×10^{-4}, -1×10^{-4}, R^2=0.24, p=0.002), and not with thalamic or cortical LV.	Fair
Mühlau et al., Mult Scler, 2013 (e73)	249 CIS (81) RRMS (168)	Cortical GM, regional volume. Deep GM, regional volume.	Global T2 lesion volume. Regional T2 lesion volume.	Global T2 LV associated significantly with lower regional GMV in the visual, primary auditory, motor, and somatosensory cortex, cerebellum, pulvinar nucleus, putamen and CN (p<0.05)[a]. In a linear multiple regression model, lower CN volume was associated with T2 lesion probability primarily in frontal WM. Lower pulvinar volume was associated with T2 lesion probability in a more widespread area, mainly in parietal and occipital WM (p<0.05)[a].	Good
Pareto et al., AJNR, 2015 (e74)	131 CIS (91) RRMS (40)	Cortical GM, regional thickness. Deep GM, regional volume.	Global T2 lesion volume.	On multivariate analyses, controlling for the presence and volume of juxtacortical lesions, significant associations were found between T2 LV and Cth in the left hemisphere (R^2=0.646, p=0.039), in particular the left parietal (R^2=0.649, p=0.035) and left occipital (R^2=0.639, p=0.038) lobules. For the right hemisphere, this was found for the right frontal (R^2=0.639, p=0.043), right parietal (R^2=0.696, p=0.020) and right occipital (R^2=0.753, p=0.007) lobules[a]. T2 LV did not reach significance in any DGM region.	Fair
Rocca et al., Radiology, 2010 (e85)	73 CIS (20) RRMS (34) SPMS (19)	Deep GM, regional volume fraction (to ICV) (thalamus).	Global T1 and T2 lesion volume.	Significant negative correlation between thalamic fraction and T2 LV (r=-0.75, p<0.001) and T1 LV (r=-0.60, p<0.001)[a].	Fair

Roosendaal et al., Mult Scler, 2011 (e1)	927 CIS (95) RRMS (657) SPMS (125) PPMS (50)	Total GM, global volume.	Global T1 and T2 lesion volumes (log-transformed for statistical analysis).	In separate multiple regression analyses, Log T1 and T2 LV were significant explanatory MRI-variables of GMV (both:β=-0.27, R^2=0.59 p<0.001)[a].	Fair
Sailer et al., Brain, 2003 (e75)	20 RRMS (11) SPMS (9)	Cortical GM, global thickness. Cortical GM, regional thickness.	Global T1 and T2 lesion volume.	Significant association between Cth and T1 [F(1,59)=6.17, p=0.03] and T2 [F(1,59)=18.04, p=0.001] LV[a]. Patients with a T1 LV of 3 ml or more showed significant cortical thinning when compared to healthy controls, mainly in frontal and temporal areas, which was not seen for patients with lower T1 LV. Patients with both low and high T2 LV had increased whole brain cortical thinning when compared to controls, but motor cortex thinning was only present in patients with a T2 LV of 20 ml or more (all p-values<0.05)[a].	Fair
Sanfilipo et al., Neuroimage, 2005 (e67)	41 RRMS (35) SPMS (6)	Total GM, global volume.	Global T1 and T2 lesion volume.	Significant correlation between global GMV and T1 LV (r=-0.46), and T2 LV (r=-0.43) (p<0.01)[a].	Fair
Sepulcre et al., Arch Neurol, 2009 (e78)	61 CIS (22) RRMS (28) SPMS (5) PPMS (6)	Cortical GM, regional volume (occipital cortex). Deep GM, regional volume (lateral geniculate nucleus).	Regional presence of T1 and Gd-enhancing lesion volume.	Significant association in the right, but not left hemisphere, between LGN volume and presence of T1 lesions in the OR (F=26.23, R^2=0.28, p=0.02)[a]. No other WM lesions, either within or outside the optic pathway, associated with LGN volume. No association between occipital CV and the presence of T1 lesions anywhere in the brain.	Fair
Shiee et al., PLoS One, 2012 (e76)	60 RRMS (43) SPMS (9) PPMS (8)	Cortical GM, global volume. Deep GM, regional volume.	Global T2 lesion volume.	Significant correlation between T2 LV and CN (r=-0.32, p<0.01) and thalamus (r=-0.36, p<0.005) volume[a], but not with putamen or global CV.	Fair
Tedeschi et al., Neurology, 2005 (e68)	597 RRMS (427) SPMS (140) PPMS (30)	Total GM, global fraction (to ICV).	Global abnormal WM fraction (aWMf) (to ICV).	Significant negative correlation between aWMf and GMF (r=-0.58, 95%CI:-0.63, -0.52), p<0.001).	Fair

Zimmerman et al., J Neural Transm, 2015 (e69)	37 RRMS (34) SPMS (3)	Total GM, global fraction (to TBV). Deep GM, regional fraction (to TBV) and volume (putamen).	Global T2 lesion volume.	Significant correlation between T2 LV and GMF (r=-0.576, p=0.016), and putamen fraction (r=-0.674, p=0.003)[a]. Patients with predominantly spinal lesions demonstrated significantly larger putamen volumes compared to patients with cerebral lesions and no detectable spinal lesions (p=0.018)[a].	Poor
Comparison between disease phenotypes					
Antulov et al., J Neuroimaging, 2011 (e23)	RRMS (67) SPMS (43)	Total GM, global volume fraction (to ICV). Cortical GM, regional volume fraction (to ICV). Deep GM, regional volume fraction (to ICV).	Regional T1 and T2 lesion volume.	Significant correlations between regional T2 LV and total GMF (controlling for regional GMF) were found in 9 and 4 of 26 regions for all MS and RRMS patients, respectively (r-values ranging from -0.20 to -0.49, $p \leq 0.001$)[a]. No significant correlations were seen for SPMS patients. Significant correlations between regional T1 LV and total GMF (controlling for regional GMF) were found in 5 regions for all MS patients, and 3 regions for RRMS patients (r-values ranging from -0.33 to -0.50, $p \leq 0.001$)[a]. No correlations were significant in the SPMS group. For all MS, RR and SPMS patients, correlations between regional T2 LV and regional GMF (controlling for total GMF) were not significant for any of the 26 regions, whereas for all MS patients, regional T1 LV approached significance with GMF in the right medial orbital frontal (r=-0.267) and left posterior basal ganglia/thalamus (r=-0.276) ($p \leq 0.01$)[a].	Fair
Bergsland et al., AJNR, 2012 (e2)	CIS (212), Early RRMS (177)	Cortical GM, global volume. Deep GM, global volume. Deep GM, regional volume.	Global T2 lesion volume and number. Global presence of Gd-enhancing lesions.	CIS patients with T2 LV above median (4.49 mL) had lower DGM (-3.97%, SED=0.51), CN (-5.73%, SED=0.13), thalamus (-4.42%, SED=0.19) (all p<0.001), GP (-2.50, SED=0.04, p<0.007), hippocampus (-3.70, SED=0.13, p<0.004), and putamen (-2.67, SED=0.14, p<0.01) volumes compared to those with T2 LV below median. CV did not differ	Good

				between CIS patients with T2 LV above or below the median value. In CIS patients, no differences in any GM volume measure were found between the subgroups when divided for the number of T2 lesions and presence of Gd-enhancing lesions. In patients with CIS, correlations were found between T2 LV and total DGM, CN, thalamus (all $p<0.0001$), and hippocampus ($p=0.001$) volumes (r-values ranging from -0.218 to -0.329)[a]. In patients with early RRMS, modest to strong correlations were found between T2 LV and every GM volume measure tested: NCV, total DGMV, CN, putamen, GP, thalamus, hippocampus, nucleus accumbens (all $p<0.0001$) and amygdala ($p=0.001$) volumes (r-values ranging from -0.246 to -0.619)[a].	
Ceccarelli et al., Neuroimage, 2008 (e40)	CIS (28) RRMS (26) SPMS (27) PPMS (18)	Cortical GM, regional volume. Deep GM, regional volume.	Global T2 lesion volume. Regional T2 lesion volume.	Significant correlation between T2 LV and lower GMV in the right (r=-0.70) and left (r=-0.81) thalamus in patients with RRMS[a]. Significant correlation between T2 LV and lower GMV in the thalami, CN, bilateral middle frontal gyrus, left inferior parietal lobule, bilateral parahippocampal gyrus and bilateral superior and inferior colliculus (r-values ranging from -0.63 to -0.88), in patients with SPMS[a]. Significant negative correlation between T2 LV and left thalamus (r=-0.94), left parahippocampal gyrus (r=-0.83) and right precentral gyrus (r=-0.85) volume in patients with PPMS[a]. In SPMS and PPMS, lower GMV in a given lobe was correlated to T2 LV within the same or adjacent lobes (r-values ranging from -0.65 to -0.91), while lower DGMV was associated to T2 LV in all lobes analysed (r-values ranging from -0.64 to -0.93)[a].	Fair

				All $p<0.001$.	
De Stefano et al., Neurology, 2003 (e31)	RRMS (65) PPMS (25)	Cortical GM, global volume.	Global T2 lesion volume.	In patients with RRMS, T2 LV correlated with CV (r=-0.47, $p<0.001$)[a]. In PPMS patients, no significant correlation was found.	Fair
Grothe et al., J Neurol, 2016 (e16)	RRMS (163) SPMS (50)	Total GM, global volume. Cortical GM, regional volume. Deep GM, regional volume. Cerebellar GM, regional volume.	Global T2 lesion volume.	Significant inverse relationship between total GMV and T2 LV in all patients (r=-0.514), patients with RRMS (r=-0.334) and SPMS (r=-0.460) (all $p<0.001$)[a]. Significant associations found between T2 LV and thalamus, various cortical and cerebellar regions in patients with RRMS (peak T-values ranging from 4.72 to 11.21), cerebellar, temporal and cingulate regions in patients with SPMS (peak T-values ranging from 5.17 to 6.17), and cortical, thalamic, cerebellar and DGM regions in the whole patient group (peak T-values ranging from 4.69 to 9.84)[a] (all $p<0.05$).	Fair
Jehna et al., Ann Neurol, 2015 (e9)	CIS (91) RRMS (69)	Cortical GM, global and regional thickness.	Regional T2 lesion volume (periventricular and non-periventricular). Regional T2 lesion volume percentage (periventricular and non-periventricular) relative to total lesion volume.	Significant correlation between PV-LL% and mean Cth in patients with RRMS (r=-0.295, p=0.015)[a], but not in patients with CIS. In the entire cohort of patients, PV-LL% did not correlate with Cth. In patients with RRMS, PV-LL correlated significantly with Cth (r=-0.293, p=0.015)[a], but the NON-PV-LL did not. In patients with CIS, neither the PV-LL nor the NON-PV-LL correlated with Cth. Vertexwise analyses: In CIS patients, PV-LL% was associated with Cth in the caudal part of the anterior cingulate cortex (p=0.0466) and the right superior frontal cortex (p=0.003). In patients with RRMS, correlations were found for bilateral precuneus, left lingual cortex, left lateral occipital gyrus, bilateral superior parietal cortices, left middle temporal cortex, right rostral part of the middle	Fair

				frontal gyrus, left superior frontal cortex and right inferior parietal gyrus (p-values ranging from 0.0002 to 0.0396)[a].	
Kalinin et al., AJNR, 2020 (e53)	71 RRMS (54) SPMS (12) PPMS (5)	Deep GM, global volume percentage (of ICV). Deep GM, total and regional volume percentage (of ICV).	Global T1 and T2 lesion volume.	All MS patients: DGMV was predicted by T1 LV, independent of T2 and cortical LV (R^2=0.52, standardized β=-0.41, p=0.02)[a]. Separated by disease duration: T2 LV was associated with DGM (R^2=0.34, standardized β=-0.70, SE=0.32, p=0.04), putamen (R^2=0.36, standardized β=-0.79, SE=0.31, p=0.02) and pallidum (R^2=0.32, standardized β=-0.77, SE=0.32, p=0.03) volume, in those with a disease duration <5 years. T1 LV was associated with DGM volume (R^2=0.71, standardized β=-0.62, SE=0.17, p=0.001) in those with a disease duration >5 years. Separated by EDSS: T2 LV was associated with thalamus volume (R^2=0.34, standardized β=-0.60, SE=0.22, p=0.01) (but none of the other volumes) in those with EDSS <4.0. No associations for T1 LV in this group. T1 LV was associated with DGM (R^2=0.77, standardized β=-0.79, SE=0.22, p=0.002) and thalamus (R^2=0.55, standardized β=-0.74, SE=0.31, p=0.03) volume in those with EDSS ≥4. No associations for T2 LV in this group. RRMS: Only T2 LV, and not T1 or cortical LV, associated with DGM (R^2=0.49, standardized β=-0.49, SE=0.22, p=0.04) and putamen (R^2=0.27, standardized β=-0.57, SE=0.27, p=0.04) volume. SPMS:	Fair

				T1 LV associated with DGMV (R^2=0.81, standardized β=-1.11, SE=0.44, p=0.04), but T2 LV did not.	
Pontillo et al., AJNR, 2019 (e55)	RRMS (52) Progressive MS (PMS) (25)	Deep GM, regional volume.	Global T2 lesion volume.	In regressions in which microstructural parameters from advanced MRI were also included, T2 LV was an independent predictor of lower volume in thalamus (standardized β=-0.347, t=-3.383, p=0.002), CN (standardized β=-0.279, t=-2.696, p=0.009), putamen (standardized β=-0.278, t=-2.367, p=0.02) and GP (standardized β=-0.481, t=-4.308, p<0.001) in all MS patients, and in thalamus (standardized β=-0.454, t=-3.214, p=0.003), CN (standardized β=-0.411, t=-2.938, p=0.006), putamen (standardized β=-0.502, t=-4.009, p<0.001) and GP (standardized β=-0.450, t=-3.407, p=0.001) in the RRMS patient group[a]. In the PMS group, this association was not found.	Fair
Riccitelli et al., HBM, 2011 (e45)	RRMS (22) SPMS (29) PPMS (22)	Cortical GM, regional volume. Deep GM, regional volume.	Local T2 lesion frequency map (Lesion Probability Map).	In RRMS and SPMS patients, a correspondence (by visual inspection) between the focal distribution of visible T2 lesions in WM structures and lower GMV in regions spatially close or known to be anatomically connected to these structures was seen. Such association was not found in PPMS patients.	Poor
Sicotte et al., Brain, 2008 (e52)	RRMS (23) SPMS (11)	Deep GM, regional volume (hippocampus and its subregions).	Global T2 lesion volume.	In RRMS patients, T2 LV was not correlated with total or subregional hippocampal volumes. In SPMS patients, there was a significant correlation between T2 LV and CA1 atrophy (r=-0.69, p=0.018)[a].	Fair
Steenwijk et al., Brain, 2016 (e34)	RRMS (130) SPMS (53) PPMS (25) Same patient group as in studies listed below (Steenwijk et al., HBM, 2015 and Steenwijk et al.,	Cortical GM, global thickness.	Global T2 lesion volume.	In a stepwise multiple regression analysis corrected for age and sex, T2 LV contributed to the model (with a negative association) for global Cth in all MS (standardized β=-0.172, p<0.001), RRMS (standardized β=-0.319, p<0.01) and PPMS (standardized β=-0.425, p<0.05)[a], but not in SPMS. For all MS and RRMS, FA in NAWM was also in the model, while for SPMS, there were no imaging metrics in the model at all.	Fair

	Radiology, 2014).				
Steenwijk et al., HBM, 2015 (e46)	RRMS (130) SPMS (53) PPMS (25) Same patient group as in studies listed over and below (Steenwijk et al., Brain, 2016 and Steenwijk et al., Radiology, 2014).	Cortical GM, regional thickness. Deep GM, regional volume.	Regional T2 lesion volume.	For the whole MS group, regional linear regression analyses showed that T2 LV in connected WM tracts explained lower DGMV in 7 out of 7 regions (avg. standardized β=-0.332), while lower Cth was explained by T2 LV in connected tracts in 28 out of 34 regions (avg. standardized β=-0.116)[a]. All regression analyses included mean tract NAWM FA as predictor. In RRMS patients, 26 of 34 models for regional Cth were significant and LV was a significant predictor in most models (avg. standardized β=-0.187). In SPMS 8 models were significant, with LV as a significant predictor in only 1 model (standardized β=-0.021). In PPMS 7 models were significant, with LV as a significant predictor in 2-3 models (avg. standardized β=-0.201)[a]. All 7 models for DGMV were significant in RRMS (LV significant predictor in 6/7, avg. standardized β=-0.349) and SPMS (LV significant predictor in 5/7, avg. standardized β=-0.264). In PPMS, all models except the models for amygdala, putamen and thalamus were significant, and LV was a significant predictor in all 4 models (avg. standardized β=-0.438)[a]. All $p<0.05$.	Fair
Steenwijk, Radiology et al., 2014 (e17)	RRMS (130) SPMS (53) PPMS (25) Same patient group as in studies listed above (Steenwijk et al., HBM,	Total GM, global volume. Cortical GM, global thickness. Deep GM, global volume.	Global T2 lesion volume.	Partial correlations with T2 LV corrected for age and sex were significant in all groups for GMV (all MS: r=-0.52, RRMS: r=-0.57, SPMS: r=-0.36, PPMS: r=-0.68, all $p<0.01$), DGMV (all MS: r=-0.70, RRMS: r=-0.74, SPMS: r=-0.65, PPMS: r=-0.71, all $p<0.001$) and Cth (all MS: r=-0.43, RRMS: r=-0.55, PPMS: r=-0.43, all $p<0.05$)[a] with as the sole exception the correlation between Cth and T2 LV in SPMS.	Fair

	2015 and Steenwijk et al., Radiology, 2016).	In multiple regression analyses, T2 LV remained a significant predictor for GMV, DGMV and Cth in the total patient group (GMV: standardized β=-0.226, DGMV: standardized β=-0.407, Cth: standardized β=-0.216, all $p<0.01$) and patients with RRMS (GMV: standardized β=-0.231, DGMV: standardized β=-0.407, Cth: standardized β=-0.357, all $p<0.01$)[a]. For patients with SPMS, T2 LV remained as a significant predictor only for DGMV (standardized β=-0.385, $p<0.001$)[a] (absent in models predicting GMV and Cth). In patients with PPMS, T2 LV was not included in any of the models (never significant).

Abbreviations: WM, white matter; GM, gray matter; DGM, deep gray matter; CIS, clinically isolated syndrome; RRMS, relapsing-remitting multiple sclerosis; SPMS, secondary progressive multiple sclerosis; PPMS, primary progressive multiple sclerosis; PMS, progressive multiple sclerosis; Gd, gadolinium; LV, lesion volume; PV-LL, periventricular lesion load; GMV, gray matter volume; DGMV, deep gray matter volume; CV, cortical volume; NCV, normalized cortical volume; ICV, intracranial volume; TBV, total brain volume; Cth, cortical thickness; CSF, cerebrospinal fluid; GMF, gray matter fraction; aWMf, abnormal white matter fraction; NAWM, normal appearing white matter; FA, fractional anisotropy; GP, globus pallidus; CN, caudate nucleus; LGN, lateral geniculate nucleus; RTV, relative thalamic volume; CA1, cornu ammonis 1; OR, optic radiation; CST, corticospinal tract; PMC, primary motor cortex; RD, radial distance; EDSS, expanded disability status scale; FDR, false discovery rate; SS, sums of deviation squares; ms, mean square; SE, standard error; SED, standard error of difference; MRI, magnetic resonance imaging; T, tesla.

[a] Full statistical details (e.g., point and/or interval estimates) not provided in published original article or supplemental material.

Supplemental eTable 2. Characteristics of longitudinal studies

Longitudinal studies					
Reference (e-ref)	*Patients (n)* *Follow-up time /(variability)*	*GM atrophy measure*	*WM lesion measure*	*Results*	*Quality of evidence*
CIS					
Dalton et al., Brain, 2004 (e6)	58 3 years (range: 2.6-6 years)	Total GM, global fraction (to ICV) percentage change.	Global T1 and T2 lesion volume and volume change. Global Gd-enhancing lesion number.	Baseline T1 and T2 LV, and number of Gd-enhancing lesions did not correlate with GMF percentage change. GMF percentage change correlated negatively with changes in T1 (r=-0.3071, p=0.0426) and T2 (r=-0.4280, p=0.0037) LV[a].	Good
Varosanec et al., AJNR, 2015 (e7)	210 4 years	Total GM, global volume percentage change. Cortical GM, global volume percentage change. Deep GM, global volume percentage change. Deep GM, regional volume percentage change.	Global T2 and Gd-enhancing lesion volume and volume percentage change. Global T2 and Gd-enhancing lesion number. Global presence of Gd-enhancing lesions.	Over the follow-up period, global GM and cortical volume change was associated with baseline T2 and Gd-enhancing LV and number, as well as the presence of Gd-enhancing lesions (all $p \leq 0.004$). DGMV change was associated with baseline Gd-enhancing lesion number (p=0.029), and thalamus volume change associated with baseline T2 LV and T2 and Gd-enhancing lesion number (all $p \leq 0.018$)[a]. Patients with the highest baseline number of T2 and Gd-enhancing lesions progressed the most in cortical (all p<0,001) and thalamic (all $p \leq 0.018$) volume change[a]. Significant associations between the total cumulative number of new/enlarging T2 lesions and global GM, cortical, DGM and thalamic volume change (p-values ranging from 0.013 to 0.036). No such associations were found for changes in T2 or Gd-enhancing LV, or cumulative Gd-enhancing lesion number[a].	Good
RRMS					

Battaglini et al., J Neurol Sci, 2009 (e50)	59 3 years (range: 2-4.8 years)	Cortical GM, regional volume change.	Regional T2 lesion volume change.	The increase in T2 LV over 3 years appeared (by visual inspection) more pronounced alongside areas with significant CV loss.	Poor
Bendfeldt et al., Neuroimage, 2009 (e27)	89 1 year (SD: ±0.9 months)	Total GM, global volume change. Cortical GM, regional volume change.	Global T1 and T2 lesion volume change. Global T2 lesion number.	Patients with increasing T1 and T2 LV after 1 year follow-up (n=45) showed significant global GM atrophy, and regional GM loss in the anterior and posterior cingulate, the temporal cortex, cerebellum, frontal and parietal cortex, the uncus and precuneus (all $p<0.01$)[a]. No GMV changes were found in patients without increasing lesion burden (n=44).	Fair
Bendfeldt et al., HBM, 2010 (e51)	89 1 year (SD: ±0.9 months)	Cortical GM, regional volume change.	Regional T1 and T2 lesion volume change.	In patients with progressive T1 and T2 LV (n=45), baseline WM lesion distribution (LPMs) did not correlate with cortical GM loss. In the same patient group, areas of significant LV changes did not anatomically overlap (by visual inspection) with areas of significant GMV loss.	Poor
Damasceno et al., Mult Scler, 2016 (e48)	42 2 years	Cortical GM, global thickness atrophy rate. Deep GM, regional volume atrophy rate.	Global T2 and Gd-enhancing lesion number. Global presence of Gd-enhancing lesions.	Patients with new/enlarging T2 or Gd-enhancing lesions had increased DGM atrophy rates (mean atrophy rate of -4.51 ±2.72 mm^3, F=4.14, p=0.024), but not thalamus or cortical atrophy rates, compared to patients without new/enlarging lesions. Significant interaction effect for group by time for Cth and thalamus volume over the 24-month period, where the group with MRI activity showed more cortical thinning (F=3.36, p=0.040) and thalamus atrophy (F=2.79, p=0.067). Similar interaction was not seen for total DGM atrophy[a]. Absence of new/enlarging T2 lesions over the follow-up was the only predictor of cortical thinning (β=1.85, 95% CI:0.61-3.08, p=0.0044), DGM (β=2.09, 95% CI:0.50-3.68, p=0.011) and thalamic atrophy rate (β=2.09, 95% CI:0.33-3.84, p=0.021).	Fair

Masuda et al., J Neurol Sci, 2019 (e29)	49 1 year	Total GM, global volume atrophy rate.	Global T2 lesion volume change.	The change in T2 LV over the follow-up did not correlate significantly with the annualized atrophy rate of the total GMV.	Fair
Preziosa et al., Neurotherapeutics, 2019 (e28)	55 2 years	Total GM, global volume percentage change.	Global T1, T2 and Gd-enhancing lesion volume change. Global T1, T2 and Gd-enhancing lesion number.	In patients treated with fingolimod (but not natalizumab), increased T2 LV correlated with percentage GMV change after two years (r=-0.43, p=0.03).	Fair
Talmage et al., PLoS One, 2017 (e57)	15 1.8 years	Deep GM, regional volume change (thalamus).	Global T2 lesion volume.	Significant correlation between baseline T2 LV and thalamic volume loss over the follow-up (r=-0.586, R^2=0.344, p=0.027)[a].	Poor
Vidal-Jordana et al., J Neuroimaging, 2016 (e26)	84 15 months	Total GM, global volume percentage change.	Global Gd-enhancing lesion number. Global presence of Gd-enhancing lesions.	GMV decreased over the follow up, regardless of the presence of Gd-enhancing lesions at baseline or at follow-up. Increased number of Gd-enhancing lesions at baseline was not predictive of GMV change.	Good
Vidal-Jordana et al., Mult Scler, 2013 (e25)	39 2 years	Total GM, global volume fraction change.	Global T2 and Gd-enhancing lesion number. Global presence of Gd-enhancing lesions.	Patients with, and without Gd-enhancing lesions at baseline, did not differ in GMF change over the follow-up period. In a linear regression analysis, no predictors were found for changes in GMF.	Fair
PPMS					
Sastre-Garriga et al., Brain, 2005 (e60)	31 1 year	Total GM, global volume fraction change.	Global T2 lesion volume. Global Gd-enhancing lesion number.	No significant associations for GMF change were found with either baseline T2 LV, or number of Gd-enhancing lesions. In stepwise regression models, baseline T2 LV and number of Gd-enhancing lesions did not predict GMF percentage change.	Fair
MS					
Pongratz et al., Brain Behav, 2019 (e87)	144 CIS (80) RRMS (64)	Total GM, global volume change.	Global T2 lesion volume change.	T2 lesion shrinking was not associated with changes in GMV, after about 1 year and 3 years of follow-up.	Fair

	3 years (range: 2.5-3.5 years)				
Fox et al., Int J Mol Sci, 2016 (e90)	92 CIS (2) RRMS (69) SPMS (16) PPMS (4) RPMS (1) 1 year	Deep GM, regional volume change (lateral geniculate nucleus).	Regional T1 lesion number.	The relative frequency of patients with volume decrease in LGN over 12 months was higher in groups with higher lesion numbers along the optic radiation. In a logistic regression analysis of LGN volume reduction and ipsilateral lesion number along the optic radiation, the number of new (OR=14.01, AUC=0.65, p=0.0006), chronic enlarging (OR=1.90, AUC=0.59, p=0.0056) and chronic shrinking (OR=12.41, AUC=0.66 p=0.0001) T1 lesions were all significant predictors[a].	Poor
Lee et al., Mult Scler, 2018 (e88)	19 RRMS (12) SPMS (7) 5.5 years (range: 1.5-10.5 years)	Total GM, global volume percentage change.	Global T1 lesion volume.	Baseline T1 LV was not a significant predictor of GMV loss over the mean follow-up duration of 5.5 years after IA/aHSCT.	Fair
Tedeschi et al., Mult Scler, 2009 (e89)	267 RRMS (208) SPMS (59) 2 years (range: 24-26 months)	Total GM, global fraction (to ICV) and percentage fraction change.	Global abnormal WM lesion fraction (to ICV) and percentage fraction change.	No significant correlation between aWMf at baseline and change in GMF during the 2 year follow up. Significant correlation between baseline GMF and change in aWMf during the 2 year follow up (r=-0.180, p=0.000)[a].	Good
Comparisons between disease phenotypes					
Fisher et al., Ann Neurol, 2008 (e24)	CIS (7) RRMS (36) SPMS (27) 4 years (range: 3.4-4.8 years)	Total GM, global volume fraction change.	Global T2 lesion volume and volume change. Global presence of Gd-enhancing lesions.	RRMS: in multiple regression models, the presence of Gd-enhancing lesions at baseline (standardized β=-0.28, p=0.04) and on-study increasing T2 LV (standardized β=-0.46, p=0.0004) were associated with greater rates of GMF percentage change over 4 years[a]. SPMS: no baseline MRI-measurements or on-study changes were significant predictors of GMF change.	Good
Tavazzi et al., J Neurol, 2020 (e10)	127 CIS (20) RRMS (85) PMS (42)	Cortical GM, global volume and volume percentage change. Deep GM, regional volume and volume	Global T2 lesion volume change.	In RRMS, atrophied T2 LV was associated with baseline thalamic volume (r=-0.384, p=0.004) and thalamic volume change (r=-0.430, p=0.004)[a]. No associations were found between atrophied T2 LV and baseline CV, or CV change. In PMS and CIS, no significant associations were found.	Fair

	5.5 years (SD: ±0.5 years)	percentage change (thalamus).	In patients with disease progression over the follow-up, atrophied T2 LV correlated significantly with baseline thalamic volume (r=-0.620, p=0.003) and thalamic volume change (r=-0.672, p=0.003)[a]. No associations were found between atrophied T2 LV and baseline CV, or CV change, or in patients without disease progression.

Abbreviations: WM, white matter; GM, gray matter; DGM, deep gray matter; CIS, clinically isolated syndrome; RRMS, relapsing-remitting multiple sclerosis; SPMS, secondary progressive multiple sclerosis; PPMS, primary progressive multiple sclerosis; PMS, progressive multiple sclerosis; RPMS, relapsing-progressive multiple sclerosis; Gd, gadolinium; LV, lesion volume; GMV, gray matter volume; DGMV, deep gray matter volume; CV, cortical volume; ICV, intracranial volume; Cth, cortical thickness; GMF, gray matter fraction; aWMf, abnormal white matter fraction; LGN, lateral geniculate nucleus; OR, odds ratio; AUC, area under the curve; CI, confidence interval; IA/aHSCT, immunoablation and autologous stem cell transplantation; MRI, magnetic resonance imaging; LPM, lesion probability map; SD, standard deviation.

[a] Full statistical details (e.g., point and/or interval estimates) not provided in published original article or supplemental material.

Supplemental eTable 3. Characteristics of cross-sectional and longitudinal studies

Cross-sectional and longitudinal studies					
Reference (e-ref)	*Patients (n)* *Follow-up time/ (variability)*	*GM atrophy measure*	*WM lesion measure*	*Results*	*Quality of evidence*
CIS					
Cacciaguerra et al., Mult Scler, 2019 (e11)	35 2 years	Deep GM, regional volume and volume change (hippocampus).	Regional T1, T2 and Gd-enhancing lesion volume and volume change.	Cross-sectional: significant correlation between ipsilateral T2 LV and lower hippocampal volume in the subiculum, bilaterally, at baseline and at month 3 (p-values ranging from 0.001-0.05), and in the CA1 subfield of the tail at month 12 and 24 (p-values ranging from 0.01-0.05)[a,b]. For T1 LV, similar results were found. T1 and T2 LV correlated with higher bilateral dentate gyrus volume at baseline and month 3 (p-values ranging from 0.05-0.001)[a,b]. Positive correlation between Gd-enhancing LV and left dentate gyrus volume at baseline and month 24 (p<0.05)[a,b]. Longitudinal: changes of LV did not correlate with hippocampal volume change.	Fair
RRMS					
Gaetano et al., Neurology, 2018 (e49)	2064 2 years	Cortical GM, global volume and volume change. Deep GM, global volume and volume change. Deep GM, regional volume and volume change (thalamus).	Global T2 lesion volume. Global Gd-enhancing lesion number.	Cross-sectional: at baseline, DGM volume was -0.258 (95% CI:-0.278, -0.239) and thalamus volume was -0.115 (95% CI:-0.124, -0.106) cm^3 smaller per cm^3 of T2 LV (all p<0.0001). Longitudinal: annual additional percentage volume loss per cm^2 of baseline T2 LV was 0.053% (95% CI:0.070, 0.037) for DGM, 0.058 (95% CI:0.076, 0.041) for thalamus and 0.052 (95% CI:0.068, 0.035) for cortical volume (all p-	Good

				<0.0001). Annual additional percentage volume loss for each additional Gd-enhancing lesion at baseline was 0.060% (95% CI:0.094, 0.025, p=0.0007) for DGM, 0.039 (95% CI:0.076, 0.002, p=0.0372) for thalamus and 0.046 (95% CI:0.082, 0.011, p=0.0102) for cortical volume. Baseline T2 LV was a significant predictor of cortical and DGM atrophy, regardless of the presence of Gd-enhancing lesions at baseline (all p<0.001)[a].	
Tiberio et al., Neurology, 2005 (e14)	20 2 years (range: 23-28 months)	Total GM, global volume fraction (to ICV) and percentage volume fraction change.	Global T1, T2, and Gd-enhancing lesion volume and volume change.	Cross-sectional: significant correlation between GMF and T2 (r=-0.438, p=0.047) and T1 (r=-0.496, p=0.022) LV[a], but not with Gd-enhancing LV. Longitudinal: changes in GMF were not correlated with Gd-enhancing LV at baseline, or changes in any LV over the follow-up.	Fair
Zivadinov et al., Mult Scler Int, 2013 (e32)	136 2 years	Cortical GM, global volume and percentage volume change.	Global T2 lesion volume and volume percentage change.	Cross-sectional: Significant correlation between baseline T2 LV and baseline CV (β=-1.39, SE=0.28, p<0.0001). Longitudinal: The percentage change in T2 LV was not significantly related to the percentage change in CV, not between baseline and year 1, year 1 and year 2, or baseline and year 2.	Fair
MS					
Ciampi et al., Mult Scler, 2017 (e64)	38 RRMS (36) SPMS (2) 3 years	Total GM, global volume fraction (to ICV) and percentage fraction change. Cortical GM, global volume and thickness. Cortical GM, regional volume and thickness. Deep GM, global volume. Deep GM, regional volume.	Global presence of Gd-enhancing lesions.	Cross-sectional: patients with or without Gd-enhancing lesions at baseline did not differ in GMF. No significant correlation between the presence of Gd-enhancing lesions at baseline and total, cortical, DGM (total and regional), brainstem and cerebellar GM volume, or with Cth (total and regional).	Fair

				Longitudinal: presence of Gd-enhancing lesions at baseline was not associated with any of the changes in GM over the 3-year study period.	
Eshaghi et al., Brain, 2018 (e71)	1214 CIS (253) RRMS (708) SPMS (128) PPMS (125) 2.41 years (SD: ±1.97 years)	Cortical GM, regional volume and volume change. Deep GM, regional volume and volume change. (Volumes used in an event-based model, to determine sequence of occurrence of atrophy)	Global T2 lesion volume and volume change.	Cross-sectional: significant association between the event-based model stage and T2 LV at baseline (standardized β =0.11, p<0.001)[a]. Longitudinal: no significant association between the rate of change in the event-based model stage over time and the rate of increase in T2 LV.	Fair
Fuchs et al., AJNR, 2018 (e86)	176 CIS (16) RRMS (114) SPMS (39) PPMS (7) 5 years (SD: ±0.7 years)	Deep GM, regional volume and volume change.	Global T2 lesion volume and volume change. Percentage tract disruption; the percentage of normative data-base derived connected tract streamlines that pass through the given lesion mask and are considered disrupted.	Cross-sectional: T2 LV was a significant predictor for DGM volume in 12 out of 14 regions (R^2 ranging from 0.108 to 0.198, p-values ranging from 0.000 to 0.019)[a]. Adding tract disruption to the model, 13 out of 14 DGM regions could be predicted (R^2 ranging from 0.108 to 0.503, all p=0.000)[a]. Controlling for T2 LV, tract disruption could only predict DGM volume in 3 out of 14 regions (R^2 ranging from 0.008 to 0.025, p-values ranging from 0.014 to 0.049)[a], but none after correction for multiple comparisons. Longitudinal: Controlling for T2 LV, tract disruption could predict DGM atrophy rate in 5 of 14 regions (R^2 ranging from 0.018 to 0.075, p-values ranging from 0.000 to 0.034)[a], but only 1 (R^2=0.075, p=0.000)[a] after correction for multiple comparisons.	Fair
Tsagkas et al., HBM, 2020 (e77)	243 RRMS (180) SPMS (51) PPMS (12) 4.36 years (SD: ±2.03)	Cortical GM, regional cortical thickness and cortical thickness change.	Global T2 lesion volume and volume change. Regional T2 lesion volume and volume change.	Cross-sectional: Global T2 LV was negatively associated with the average Cth in regions symmetrically in nearly the whole cortex (mean t-values =-4.00±1.36 (right hemisphere) and -4.24±1.35 (left hemisphere), all p<0.05). Similarly, regional T2 LV in the left and right frontal, temporal, parietal, and occipital areas were associated with lower Cth in widespread bilateral cortical regions (all p<0.05)[a].	Good

		Longitudinal: Changes in global T2 LV did not associate with changes in Cth over time. Left temporal changes in T2 LV associated negatively with change in Cth in the left cuneus (mean t-value=-4.47±0.35) and precuneus (mean t-value=-4.43±0.21) (all $p<0.05$). A negative association was also shown between changes in left occipital T2 LV and Cth change in bilateral temporal, parietal and occipital regions (mean t-values ranging from -3.06 to -3.97±0.09-0.54, all $p<0.05$).

Abbreviations: WM, white matter; GM, gray matter; DGM, deep gray matter; CIS, clinically isolated syndrome; RRMS, relapsing-remitting multiple sclerosis; SPMS, secondary progressive multiple sclerosis; PPMS, primary progressive multiple sclerosis; Gd, gadolinium; LV, lesion volume; DGM, deep gray matter; CV, cortical volume; ICV, intracranial volume; Cth, cortical thickness; GMF, gray matter fraction; CA1, cornu ammonis 1; SD, standard deviation; SE, standard error.

[a] Full statistical details (e.g., point and/or interval estimates) not provided in published original article or supplemental material.

[b] Spearmans rho ranges indicated in color bar in Figure 4 in the original study.

II

European Radiology (2022) 32:3576–3587
https://doi.org/10.1007/s00330-021-08405-8

NEURO

The effect of gadolinium-based contrast-agents on automated brain atrophy measurements by FreeSurfer in patients with multiple sclerosis

Ingrid Anne Lie[1] · Emma Kerklingh[2] · Kristin Wesnes[1,3,4] · David R. van Nederpelt[2] · Iman Brouwer[2] · Øivind Torkildsen[1,3] · Kjell-Morten Myhr[1,3] · Frederik Barkhof[2,5] · Lars Bø[1,6] · Hugo Vrenken[2]

Received: 11 June 2021 / Revised: 7 September 2021 / Accepted: 12 October 2021 / Published online: 3 January 2022

Abstract

Objective To determine whether reliable brain atrophy measures can be obtained from post-contrast 3D T1-weighted images in patients with multiple sclerosis (MS) using FreeSurfer.

Methods Twenty-two patients with MS were included, in which 3D T1-weighted MR images were obtained during the same scanner visit, with the same acquisition protocol, before and after administration of gadolinium-based contrast agents (GBCAs). Two FreeSurfer versions (v.6.0.1 and v.7.1.1.) were applied to calculate grey matter (GM) and white matter (WM) volumes and global and regional cortical thickness. The consistency between measures obtained in pre- and post-contrast images was assessed by intra-class correlation coefficient (ICC), the difference was investigated by paired t-tests, and the mean percentage increase or decrease was calculated for total WM and GM matter volume, total deep GM and thalamus volume, and mean cortical thickness.

Results Good to excellent reliability was found between all investigated measures, with ICC ranging from 0.926 to 0.996, all p values < 0.001. GM volumes and cortical thickness measurements were significantly higher in post-contrast images by 3.1 to 17.4%, while total WM volume decreased significantly by 1.7% (all p values < 0.001).

Conclusion The consistency between values obtained from pre- and post-contrast images was excellent, suggesting it may be possible to extract reliable brain atrophy measurements from T1-weighted images acquired after administration of GBCAs, using FreeSurfer. However, absolute values were systematically different between pre- and post-contrast images, meaning that such images should not be compared directly. Potential systematic effects, possibly dependent on GBCA dose or the delay time after contrast injection, should be investigated.

Trial registration Clinical trials.gov. identifier: NCT00360906.

Key Points

- *The influence of gadolinium-based contrast agents (GBCAs) on atrophy measurements is still largely unknown and challenges the use of a considerable source of historical and prospective real-world data.*
- *In 22 patients with multiple sclerosis, the consistency between brain atrophy measurements obtained from pre- and post-contrast images was excellent, suggesting it may be possible to extract reliable atrophy measurements in T1-weighted images acquired after administration of GBCAs, using FreeSurfer.*
- *Absolute values were systematically different between pre- and post-contrast images, meaning that such images should not be compared directly, and measurements extracted from certain regions (e.g., the temporal pole) should be interpreted with caution.*

✉ Ingrid Anne Lie
Ingrid.lie@uib.noAuthors and Affiliations

[1] Department of Clinical Medicine, University of Bergen, Bergen, Norway

[2] Department of Radiology and Nuclear Medicine, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, Location VUmc, De Boelelaan 1118, 1081 HZ Amsterdam, The Netherlands

[3] Neuro-SysMed, Department of Neurology, Haukeland University Hospital, Bergen, Norway

[4] Department of Neurology, St. Olav's University Hospital, Trondheim, Norway

[5] Institutes of Neurology and Healthcare Engineering, UCL London, London, UK

[6] Norwegian Multiple Sclerosis Competence Centre, Department of Neurology, Haukeland University Hospital, Bergen, Norway

Keywords Multiple sclerosis · Magnetic resonance imaging · Gadolinium · Atrophy · Grey matter

Abbreviations

BBB	Blood–brain barrier
CNR	Contrast-to-noise ratio
CR	Contrast ratio
EDSS	Expanded Disability Status Scale
FFE	Fast field echo.
FLAIR	Fluid-attenuated inversion recovery
FLIRT	FMRIB´s Linear Image Registration Tool
FSL	FMRIB Software Library
GBCAs	Gadolinium-based contrast-agents
GM	Grey matter
ICC	Intra-class correlation coefficient
IQMs	Image quality metrics
LST	Lesion segmentation tool
mL	Millilitres
mm	Millimetre
MPRAGE	Magnetization-prepared rapid gradient-echo
MRI	Magnetic resonance imaging
MRIQC	MRI Quality Control Tool
ms	Millisecond
MS	Multiple sclerosis
RRMS	Relapsing–remitting MS
SD	Standard deviation
SNR	Signal-to-noise ratio
SPMS	Secondary progressive multiple sclerosis
SPSS	Statistical Product and Service Solutions
T	Tesla
T1	Inversion time
TE	Echo time
TR	Repetition time
WM	White matter

Introduction

Grey matter (GM) atrophy measured on MRI in persons with multiple sclerosis (MS) reflects irreversible neuroaxonal loss and neurodegenerative changes in the CNS [1]. The degree of GM atrophy has been shown to consistently correlate with physical [2, 3] and cognitive [4] disability, and is regarded as a promising neurodegenerative biomarker. Furthermore, as the demand for neuroprotective interventions increases, GM atrophy is an easily available outcome measure [5–7].

There are a number of available methods and software to measure GM atrophy. Although FreeSurfer requires substantial processing time, making it less suitable for clinical practice, it is one of the most commonly used automated methods in research, especially for cortical parcellation and thickness estimation. FreeSurfer is publicly available and widely validated [8–12], and in the body of literature on GM atrophy in MS, many key papers have used FreeSurfer [13–15].

Due to the high tissue contrast [16–18] in unenhanced three-dimensional (3D) T1-weighted images, this image type is commonly used by brain segmentation software and required by FreeSurfer, as well as other software with similar purposes. However, unenhanced 3D T1-weighted images are not mandatory in suggested standardised brain MRI protocols for MS [19] and may not be routinely included. Instead, post-contrast T1-weighted images are often prioritised, especially in clinical settings. In case of ongoing inflammation, the intravenously administered contrast agent leaks into the brain parenchyma in locations where the blood–brain barrier (BBB) is disrupted [20]. These post-contrast images are valuable both in baseline and follow-up examinations, as they can unequivocally detect lesions with active inflammation [19].

The contrast agent used is almost universally gadolinium-based, consisting of a central paramagnetic Gd^{3+} ion chelated to a carrier molecule to prevent the toxicity of free Gd^{3+}, while still maintaining its paramagnetic properties. Gadolinium-based contrast agents (GBCAs) shorten both the longitudinal (T1) and transverse (T2) relaxation times [21], leaving areas in which GBCAs accumulate as bright or hyperintense compared to surrounding tissue on T1-weighted images.

The use of GBCAs has increased over the last three decades [22], making up a considerable source of historical and prospective real-world data. However, the value of such data for brain atrophy measurements depends on our ability to correctly interpret the data in automated image analyses. The influence of GBCAs on atrophy measurements is still largely unknown and has previously been investigated in only a few studies using different image analysis techniques [23, 24]. In this study, our aim is to validate the use of post-contrast T1-weighted images for volume and cortical thickness measurements and to provide guidelines on how to interpret results from clinically relevant and commonly considered measures. To do so, total WM and GM volume, total deep GM and thalamus volume, and mean cortical thickness measures were obtained in pre- and post-contrast images by FreeSurfer and compared.

Materials and methods

Participants

The patients included in this study participated in a 10-year follow-up visit following a multi-centre, randomised,

placebo-controlled trial of ω-3 fatty acids in MS (the OFAMS-study), which has previously been described in detail [25]. A total of 85 of the 92 persons with relapsing–remitting MS (RRMS) [26] originally enrolled in the OFAMS-study participated in the 10-year follow-up visit and underwent clinical, biochemical, and radiological examinations at their local study site.

The study was approved by the Regional Committee for Medical and Health Research Ethics in Western Norway Regional Health Authority (clinical trials.gov, identifier: NCT00360906). All participants gave their written informed consent.

MRI data and analysis

MRI data acquisition

Imaging at the 10-year follow-up visit was performed at the different study sites, on a 3-Tesla (T) MRI scanner if available, alternatively using a 1.5 T MRI scanner, with a standard head coil. The acquisition included a post-contrast sagittal 3D T1-weighted sequence; acquisition details across sites are provided in Table 1. Furthermore, a sagittal T2-weighted 3D fluid-attenuated inversion recovery (FLAIR) sequence was acquired according to locally optimised protocols. The full MRI protocol provided to the study sites is available in eAppendix 1. The study sites were encouraged to include the same 3D T1-weighted sequence before contrast-agent administration, if possible. For the present study, only the subset of the participants who underwent 3D T1-weighted MR imaging both before and after injection of GBCAs, during the same scanner visit, and with the exact same acquisition protocol, was included.

MRI data processing

Lesion segmentation and lesion filling Lesion segmentation was done on FLAIR images using lesion segmentation tool (LST) (version 2.0.15; http://applied-statistics.de/lst.html) [27]. The lesion probability map in FLAIR space was brought to T1-weighted space by FLIRT linear registration of the FLAIR image to the T1 image, using 7 degrees of freedom, correlation ratio as the cost function, and trilinear interpolation. Afterwards, a threshold of 0.1 was used to binarise the lesion probability map. To optimize the lesion filling, gadolinium-enhancing regions (both lesions and other regions) were first removed, by applying an upper-intensity threshold at the 98th percentile. Next, the FMRIB Software Library (FSL) (version 5.0.10; http://www.fmrib.ox.ac.uk/fsl) was used to fill in lesional voxels in the T1-weighted images using the lesion_filling tool [28], and these filled lesions were pasted into the original post-contrast 3D T1-weighted images.

Morphological reconstruction Cortical reconstruction and parcellation for cortical volume and thickness measurement and subcortical segmentation were performed with FreeSurfer, a freely available software package for academic use, available through online download (http://surfer.nmr.mgh.harvard.edu/). The findings presented here were obtained using FreeSurfer version 7.1.1; highly comparable findings obtained using FreeSurfer version 6.0.1 are presented in Table e1. The technical details of FreeSurfer procedures have been previously described [29, 30] and briefly summarised in eAppendix 2.

Quality control was performed by visual inspection, and any segmentation errors were recorded for each patient. In cases where only specific anatomical regions were incorrectly

Table 1 Details on MRI acquisition per protocol

Protocol (number of patients)	1 (3)	2 (3)	3 (3)	4 (3)	5 (5)	6 (2)	**7** (3)
Scanner	Siemens Aera	Siemens Skyra	Siemens Avanto	Siemens Aera	Philips Achieva	Siemens Prisma	Philips Achieva
Field strength	1.5 T	3 T	1.5 T	1.5 T	1.5 T	3 T	1.5 T
3DT1 sequences	MPRAGE	MPRAGE	MPRAGE	MPRAGE	FFE	MPRAGE	FFE
TR (ms)	1940	2300	2200	2200	7.6	1800	7.1
TE (ms)	2.69	2.32	2.82	2.67	3.75	2.28	2.2
TI (ms)	976	900	900	900		900	
Flip angle (°)	8	8	8	8	8	8	8
Voxel size	1.00×0.98×0.98	0.9×0.94×0.94	1.00×0.98×0.98	1.00×0.98×0.98	1.00×0.98×0.98	1.00×0.50×0.50	1.00×1.00 ×1.00
Head receiver coil	Unknown	Unknown	HE1-4	HE1-4	SENSE-head-8	Unknown	SENSE-head-8
GBCA	Gadoterate meglumine	Gadoterate meglumine	Gadoterate meglumine	Gadoteridol	Unknown	Gadoterate meglumine	Unknown

Abbreviations: *TR*, repetition time; *TE*, echo time; *TI*, inversion time; *ms*, millisecond; *mm*, millimetre; *MPRAGE*, magnetization-prepared rapid gradient-echo; *FFE*, fast field echo; *GBCA*, gadolinium-based contrast agent

segmented, we chose to not apply any corrections for these errors in our analyses.

The Desikan-Killiany atlas [31] was used to extract cortical thickness measures (mean cortical thickness, left and right hemisphere) and to study regional differences in cortical thickness between pre- and post-contrast images, across subjects, by creating a heat map. Furthermore, total cerebral GM and WM volume and total deep GM and thalamus volume (left and right hemisphere) were obtained.

MRI quality control tool To investigate potential root causes of any observed segmentation differences, both pre- and post-contrast T1-weighted images were analysed using the MRI Quality Control Tool (MRIQC) [32]. MRIQC is an open-source software and extracts no-reference image quality metrics (IQMs) from structural and functional MRI data [32]. Using a segmentation into GM, WM, and CSF by FSL-FAST [33], MRIQC calculates tissue-specific signal-to-noise ratio (SNR) values as well as the contrast-to-noise ratio (CNR) between GM and WM. Additionally, based on these values obtained from MRIQC, the contrast ratio (CR) between white and grey matter was also calculated.

Statistical analysis

Statistical analyses were performed using the Statistical Product and Service Solutions (SPSS) for macOS (Version 25; SPSS). Data were visually and statistically examined using the Kolmogorov–Smirnov test for normality. To assess the agreement between volume and thickness measurements obtained before and after GBCA administration, the intra-class correlation coefficient (ICC) was determined, based on a mean rating ($k=2$), consistency, two-way mixed model. Scatterplots were created to visualise the agreement. To assess whether any systematic differences in structural measurements or IQMs were present between pre- and post-contrast measurements, paired t-tests were performed. Furthermore, boxplots were made to illustrate any differences, and Bland–Altman plots were created to identify fixed or proportional bias [34]. As an exploratory analysis, paired t-tests were used to investigate a possible systematic difference between field strengths (1.5 and 3 T).

Results

Pre- and post-contrast T1-weighted images were obtained with the exact same acquisition protocol in a total of 23 patients. One patient was excluded due to a large image artifact, causing segmentation errors. Table 2 provides an overview of the demographic and clinical characteristics of the patient group.

Table 2 Demographic and clinical characteristics

Age in years, mean (SD)	50.5 (7.83)
Sex, female, *N* (%)	15 (68.2%)
Disease duration, mean in years (SD) / median (range)	13.8 (3) / 13 (12–25)
EDSS, mean (SD) / median (range)	2.9 (1.2) / 2.5 (1–6)
Disease phenotype (N)	RRMS (21), SPMS (1)
Study site (number of patients)	Site 1 (3), Site 2 (3), Site 3 (3), Site 4 (8 (2 scanners)), Site 5 (2), Site 6 (3)

Abbreviations: *SD*, standard deviation; *EDSS*, Expanded Disability Status Scale; *RRMS*, relapsing-remitting multiple sclerosis; *SPMS*, secondary progressive multiple sclerosis

Quality control of FreeSurfer segmentations

All 22 pairs of pre- and post-contrast T1-weighted images finished the fully automated FreeSurfer pipeline (i.e., no hard failures). The most common soft failures (i.e., failures that do not disrupt the pipeline, but may need modification) are summarised in Table 3.

Volume and cortical thickness measurements before and after administration of GBCAs

The mean values of MRI measurements obtained before and after GBCAs are summarised in Table 4 and Fig. 4. Briefly, a mean increase in GM volumes and cortical thickness measures were observed in post-contrast images, while a mean decrease was observed in total WM volume. The results of the exploratory analysis subdivided according to field strength are presented in Table e2, showing no clear systematic differences between field strengths.

Consistency of measurements obtained before and after administration of GBCAs

A high degree of reliability was found between the measurements obtained pre- and post-contrast, for all volumes and cortical thickness measures assessed. All ICC values (Table 4) were above 0.92, with the lowest values in the thalami, and above 0.96 for all larger structures, all p values < 0.001. The consistency between the measurements is demonstrated in Fig. 5.

Difference in measurements before and after administration of GBCAs

GM volumes and mean cortical thickness were significantly higher after administration of GBCAs, in all investigated structures (Table 4, Figs. 4 and 5).

Table 3 Summary of the most common soft failures

Description	Frequency
The pial surface (representing the border between cortical GM and CSF) or the border of segmented deep GM structures, expanding into extraparenchymal tissue, including components of dura or blood vessels as part of the cortex or deep GM structures (Figs. 1b and 2)	Found in all scans, both pre- and post-contrast, but more frequently and to a more severe degree in post-contrast images
The pial surface failing to follow the white surface, causing "looping" errors (Fig. 1a) and subsequent incorrect enlargement of the cortical volume and thickness	Found in all scans, both pre- and post-contrast, but more frequently and to a more severe degree in post-contrast images
The constructed surface border between WM and GM (the white surface) failing to follow the intensity gradient correctly in the temporal poles, resulting in a suboptimal segmentation (Fig. 3)	Found to a moderate degree in two post-contrast images, and to a minor degree in a total of eight patients, in the post-contrast image in all eight, and in the pre-contrast image in three of those eight

Abbreviations: *GM*, grey matter; *CSF*, cerebrospinal fluid; *WM*, white matter

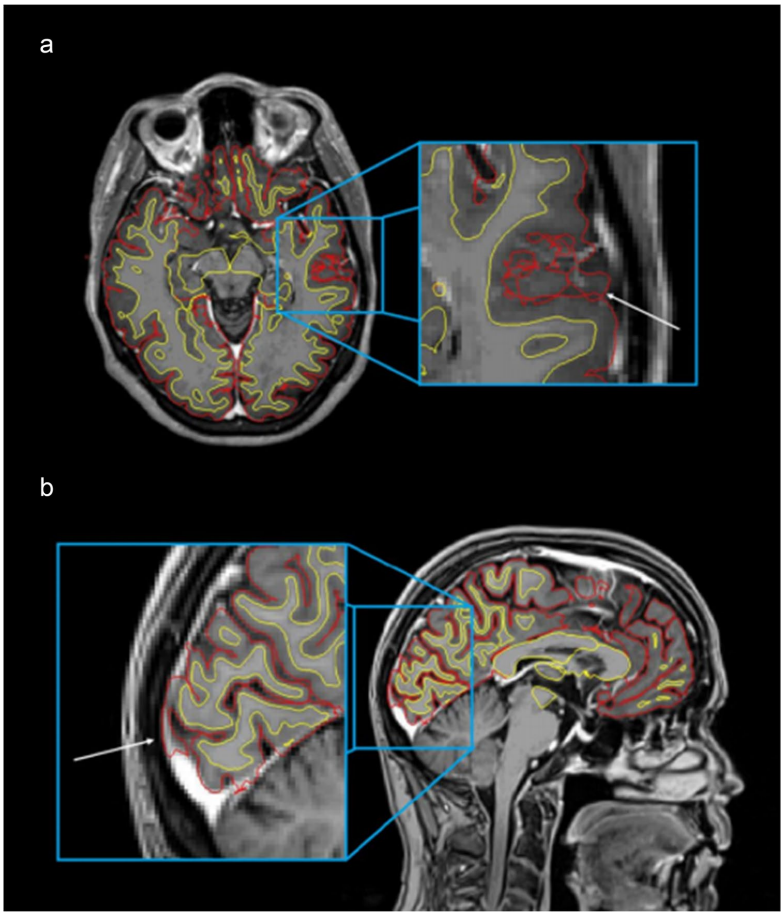


Fig. 1 Post-contrast T1-weighted MRI, showing the border between WM and GM (white surface) (yellow), and the border between GM and CSF (pial surface) (red). **a** Axial slice showing a moderate pial surface "looping error" (white arrow). **b** Sagittal slice showing a typical skull stripping failure; a moderate error of the pial surface expanding into the dura and the sagittal sinus (white arrow)

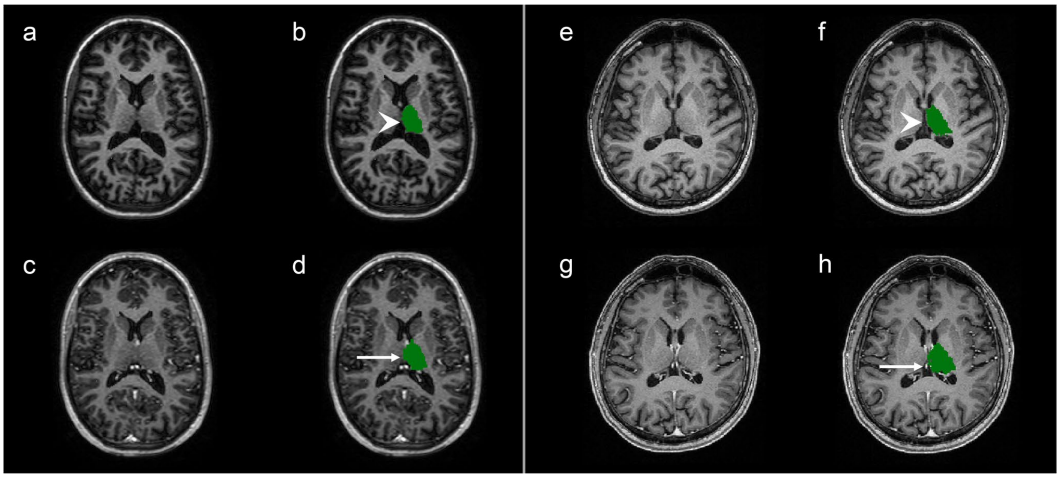


Fig. 2 T1-weighted MRI, showing the segmentation of the left Thalamus in pre- and post-contrast images, in two different patients (subject E3 (**a**–**d**) and subject C1 (**e**–**h**)). **a–d** Axial slices demonstrating the typical quality of thalamus segmentations. In post-contrast images (**c**–**d**), the medial border of the left Thalamus is slightly overestimated (arrow) compared to pre-contrast images (arrowhead) (**a**–**b**), most likely due to hyperintense signal from extraparenchymal structures in the midline. **e**–**h** Axial slices demonstrating a more severe overestimation of the medial border of the left Thalamus (arrow) in post-contrast images (**g**–**h**) compared to pre-contrast images (arrow head) (**e**–**f**). Again, the segmentation of the medial border is overestimated due to inclusion of extraparenchymal hyperintense structures, in this case, the internal cerebral vein)

Figure 6 shows heatmaps visualising the difference in cortical thickness between pre- and post-contrast images, demonstrating the general increase in thickness measured in post-contrast T1-weighted images. However, in a few exceptions, most prominently the temporal pole, the parahippocampal, and the entorhinal gyrus in the temporal lobe, cortical thickness decreased.

While GM volumes and cortical thickness measurements were higher after administration of GBCAs, total WM volume was significantly lower. Figure e1 in the supplementary material shows the constructed Bland–Altman plots, revealing systematic differences, but no proportional bias.

IQMs are reported in Table 5. The CNR was not significantly different between pre- and post-contrast images.

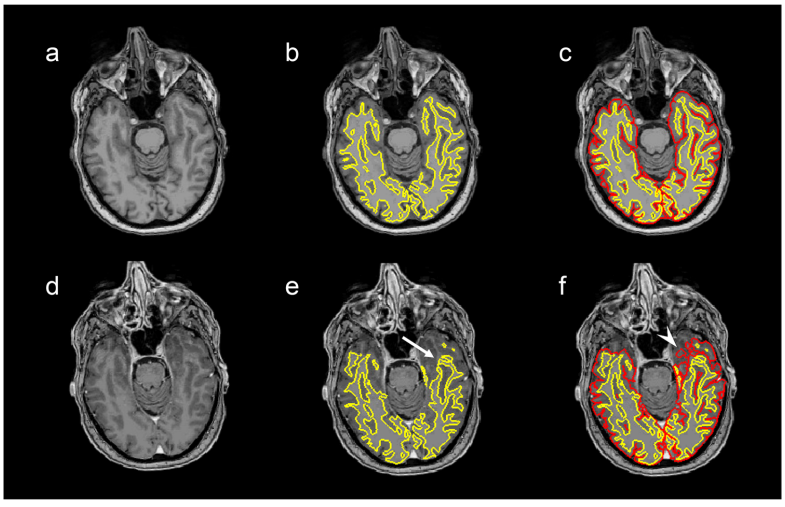


Fig. 3 Pre-contrast (**a**–**c**) and post-contrast (**d**–**f**) T1-weighted images obtained from the same patient (subject A3) in the same MRI session. **b** and **e** show the white surface, which is the border between white and grey matter as automatically constructed by FreeSurfer (yellow). **c** and **f** show the pial surface, which is similarly the automatically constructed border between grey matter and cerebrospinal fluid (red), derived from the white surface. The figure demonstrates a typical failure of moderate degree, where the white surface fails to include parts of the temporal poles in the post-contrast image (**e**) (arrow), with subsequent mistakes in the pial surface (**f**) (arrowhead)

Table 4 MRI measurement values

MRI measure	Mean value pre-contrast (SD)	Mean value post-contrast (SD)	Mean difference[a] (SD)	Percent increase/ decrease (SD)	ICC (95% confidence interval)
Total grey matter volume (mL)	602.53 (62.42)	620.33 (59.97)	17.80 (16.20)**	+3.06 (2.79) %	0.982 (0.957–0.993)
Total white matter volume (mL)	457.06 (63.05)	448.70 (59.25)	−8.36 (7.35)**	−1.74 (1.48) %	0.996 (0.991, 0.998)
Total deep grey matter volume (mL)	51.57 (5.90)	54.73 (5.61)	3.16 (2.04)**	+6.33 (4.43) %	0.968 (0.922–0.987)
Left thalamus volume (mL)	6.47 (0.93)	7.58 (1.09)	1.10 (0.48)**	+17.39 (8.46) %	0.940 (0.855–0.975)
Right thalamus volume (mL)	6.37 (0.99)	7.12 (0.92)	0.75 (0.50)**	+12.52 (9.04) %	0.926 (0.823–0.969)
Mean cortical thickness left hemisphere (mm)	2.32 (0.16)	2.49 (0.15)	0.17 (0.06)**	+7.38 (2.78) %	0.964 (0.913, 0.985)
Mean cortical thickness right hemisphere (mm)	2.33 (0.14)	2.49 (0.14)	0.16 (0.05)**	+7.13 (2.61) %	0.961 (0.906, 0.984)

Abbreviations: *SD*, standard deviation; *ICC*, intra-class correlation coefficient for consistency; *mL*, millilitres; *mm*, millimetre

[a] Paired t-test

** $p < 0.001$

Tissue-specific SNRs were significantly lower in post-contrast images, for both GM ($p < 0.01$) and WM ($p < 0.0001$). The CR between WM and GM was significantly higher in post-contrast images ($p < 0.006$).

Discussion

Our results demonstrate that using FreeSurfer, reliable GM volume- and cortical thickness measurements may be

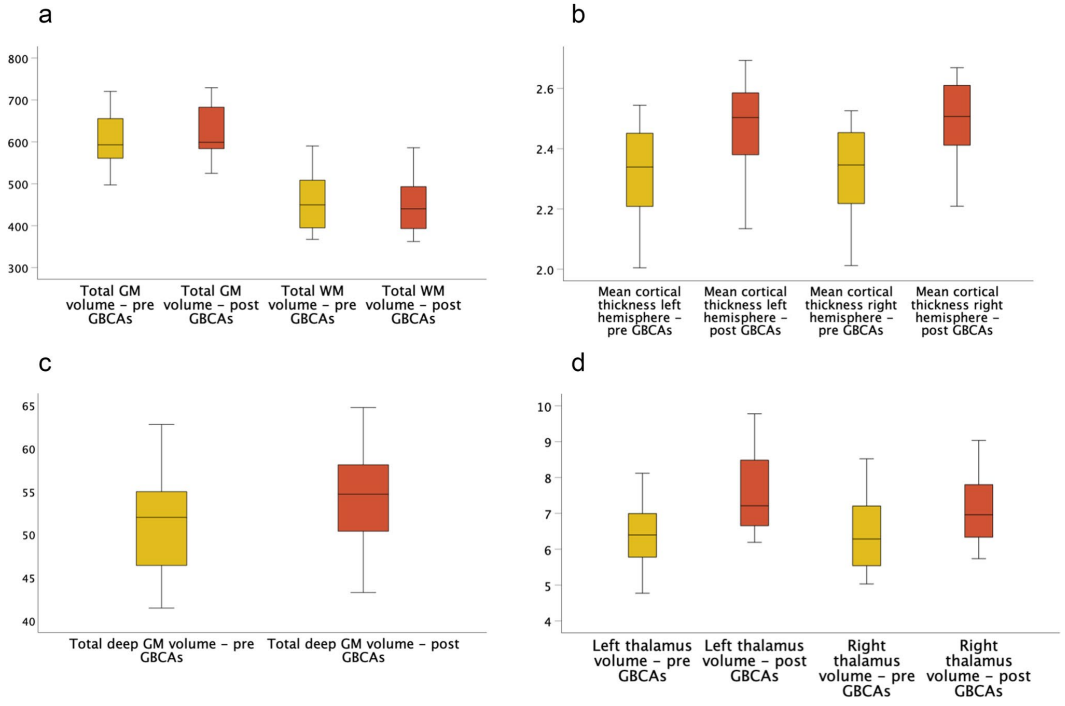


Fig. 4 Boxplots of MRI measurements obtained before (yellow) and after (red) GBCA administration, in mL (**a**, **c**, **d**) and mm (**b**)

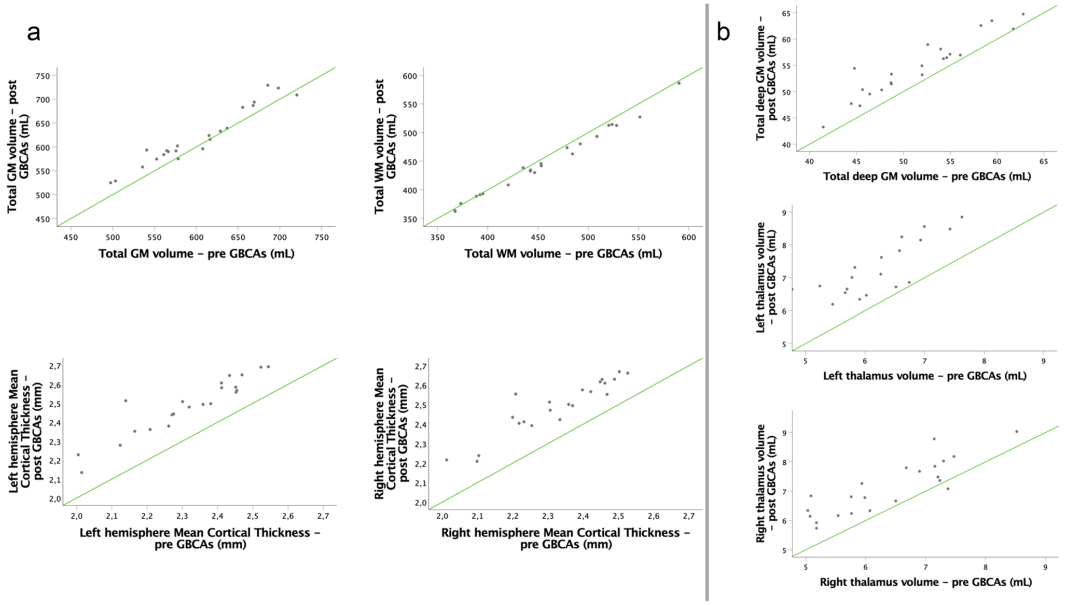


Fig. 5 Scatterplots of global (**a**) and regional (**b**) MRI measurements obtained before and after GBCA administration. The green lines indicate identity lines

obtained from post-contrast 3D T1-weighted images. Despite systematic overestimation of the GM, high consistency was observed between all investigated MRI brain measurements obtained before and after administration of GBCAs.

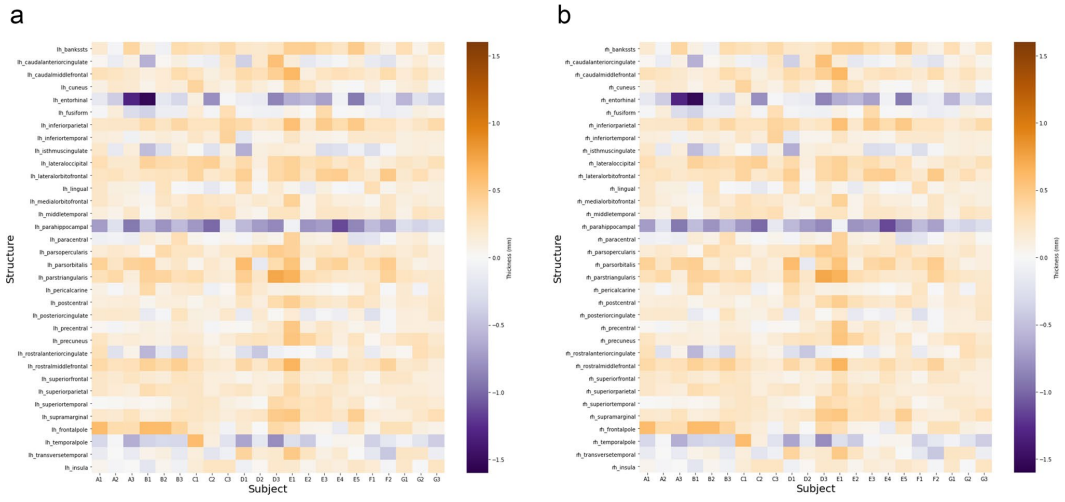


Fig. 6 Heatmaps demonstrating the difference (mm) in cortical thickness in the left (**a**) and right (**b**) hemisphere after administration of GBCAs. Brown colours indicate an increase in cortical thickness, and purple colours indicate a decrease in cortical thickness (colour range between -1.6 mm and + 1.6 mm cortical thickness difference). Letters in subject names indicate MRI scanner (**a**–**g**)

Table 5 Image quality metrics obtained by MRI Quality Control Tool

Image quality metric	Mean value pre-contrast (SD)	Mean value post-contrast (SD)	Mean difference[a] (SD)
CNR	3.23 (0.56)	3.25 (0.40)	−0.1 (0.52)
SNR GM	10.89 (2.81)	9.02 (2.50)	1.87 (2.14)*
SNR WM	18.01 (3.72)	16.24 (3.56)	1.77 (1.64)***
CR	0.39 (0.08)	0.45 (0.08)	−0.05 (0.08)*

Abbreviations: *SD*, standard deviation; *CNR*, contrast to noise ratio; *SNR*, signal to noise ratio; *GM*, grey matter; *WM*, white matter; *CR*, contrast ratio

[a] Paired t-test

*$p < 0.01$, ***$p < 0.0001$

To our knowledge, this is one of the very few studies investigating the effect of GBCAs on volume measures in MS patients and the first using FreeSurfer. In our study, when investigating the consistency between the measures obtained before and after administration of GBCAs, a good to excellent [35] reliability was found between all investigated measures. This is in agreement with previous studies investigating the whole brain [36], upper cervical cord area [37], and GM and WM measurements [23] using SIENAX [23, 36], volBrain, and FSL-Anat [23] and may imply that reliable atrophy measurements acquired from post-contrast images are possible across segmentation techniques.

Consistently, total GM, deep GM, and thalamic volume were between 3.06 and 17.39% higher in post-contrast images, and the same tendency was found for mean cortical thickness. Simultaneously, total WM volume was 1.74% lower in post-contrast images. The differences were systematic across all investigated measurements and exhibited no proportional bias. Inspecting cortical segmentations in more detail, we produced heatmaps highlighting within-subject cortical thickness differences in smaller cortical regions (Fig. 6). While smaller regions almost inevitably produce more variability than the larger regions that were the main focus of this work, these inspections showed that cortical thickness overestimation was a brain-wide phenomenon and that the overestimation in post-contrast images was not tied to large errors in any specific region but instead occurred throughout the brain.

These systematic differences in measured volumes and cortical thicknesses between pre- and post-contrast images mean that they should not be compared directly. Another study, using synthetic tissue mapping to measure brain tissue fractions [24], found a 1.1% increase in total WM fraction and an 0.7% decrease in GM fraction, in post-contrast images. Due to the methodological differences between that study and ours, it is difficult to assess the reason for the discrepancy in findings.

We could not identify any definite reason for the differences between pre- and post-contrast images. However, when visually inspecting images separately, some recurring soft failures in the FreeSurfer pipeline were found: First, the pial surface often expanded into extraparenchymal tissue, including components of dura or blood vessels as part of the cortex (Fig. 1b). These errors have been shown in areas where the dura or other structures like venous sinuses, lie tangentially in close proximity to the cortex or deep GM structures, leading to larger thickness and volume variability (Fig. 2) [38]. In the FreeSurfer processing stream, the failure to remove enough extraparenchymal tissue happens in the preliminary skull stripping step [39] and the accuracy of the pial surface can be improved by manually erasing the incorporated dura or blood vessels before rerunning analyses [40].

Another recurring soft failure concerned the pial surface. In the surface-based cortical reconstruction, the border between white and grey matter (the white surface) is delineated, following T1 intensity gradients. The pial surface is then grown from the white surface, which serves as a reference point [41]. In all images, but more frequently and severely in post-contrast images, the pial surface failed to follow the white surface, causing "looping" errors (Fig. 1a) and a subsequent incorrect enlargement of the cortical volume and thickness. To improve pial surface accuracy, it is recommended to check for any mistakes in the white surface, and possibly apply manual edits before rerunning analyses [40].

Although most cortical regions demonstrated an increase in cortical thickness in post-contrast images, there were a few exceptions, particularly in the medial part of the temporal lobe. In the entorhinal and parahippocampal gyrus, as well as in the temporal poles, the measured cortical thickness was in some patients thinner after GBCA administration. These regions have in common that they are relatively small and structurally complex, and on visual inspection of the errors, the constructed white surface did not correctly follow the intensity gradients, causing considerable errors in the white surface, and subsequently the pial surface, leaving out parts of the temporal pole (Fig. 3). Challenges in reconstructing parts of the temporal cortex are consistent with previous studies [31, 38, 40, 42], leading to increased variability of the local cortical thickness measurements [38].

The soft failures in the FreeSurfer pipeline occurred more often in post-contrast images in our data. This may be caused by the higher intensity in extraparenchymal structures in close proximity to the cortex or subcortical structures, causing disturbance and challenges in correctly separating different tissue types.

Skull stripping errors and other soft failures could in some selected regions be identified as the direct cause of increased cortical thickness or GM volume in post-contrast images. It is however uncertain if these errors can explain the systematic increase in almost all GM structures and the overall decrease in WM volume. Even in the absence of active lesions and GBCA leakage through disruptions in the BBB, GBCAs can still be expected to be present in the brain capillary network [24]. This presence may shorten the overall T1 relaxation time in all tissues, and possibly also affect intensity borders. In our MRIQC analyses, there was no difference between pre- and post-contrast images CNR, indicating that the separation of GM and WM tissue distributions was similar in pre- and post-contrast images. It should however be noted that extracting reliable noise estimates from parallel imaging is challenging.

Systematic effects dependent on the type of GBCA used, dosage, and delay time after administration are likely. In the data retrospectively used in the present study, these factors were not standardised, nor always stated, making them difficult to correct for. To further conclude on the reliability of post-contrast measurements, it is necessary for future research to investigate the possible systematic effects dependent on these variables.

This study is not without limitations. For a multicentre study, the number of patients included was limited, and patients were scanned on different scanners with varying sequence parameters and field strength. Furthermore, some details of the MRI protocol that may affect brain measurements (e.g., head coils [43, 44]) were in some cases neither stated nor retrospectively retrievable, making it difficult to evaluate the effect of these factors. Nonetheless, because the effect of field strength on atrophy measures has been studied before [45], we explored the results for 1.5 T and 3 T scanners separately. No systematic differences between the two field strengths regarding the different variabilities in the pre- and post-contrast images emerged, which could be due to small patient numbers and variable acquisition settings. Considering all these aspects, the fact that consistency between measurements before and after GBCA administration was observed across the different scanners, suggests that this behaviour is largely systematic. Future studies should investigate the effect of field strength and those of other aspects of image acquisition more systematically. Image analyses in this study were performed by FreeSurfer, while there are multiple other software packages that have been used in the MS literature. To focus the present work, we chose FreeSurfer because it allows both volumetric and cortical thickness measurements and has been widely used before in MS [46–50].

Finally, we did not perform any pre-processing to remove high-intensity regions (except for those in WM lesions, filled in the lesion filling process) from the post-gadolinium T1-weighted images. Future work should investigate whether removal or replacement of those regions, perhaps similar to the procedure followed as part of the lesion-filling process in the present work, could reduce the observed overestimation of grey matter.

Conclusion

This study has demonstrated that reliable atrophy measurements may be obtained by FreeSurfer from post-contrast 3D T1-weighted images. A good to excellent consistency was observed between all investigated GM and WM measurements derived from images acquired before and after GBCA administration. However, due to the systematic overestimation of the GM in post-contrast images, measurements acquired from pre- and post-contrast images should not be compared directly, and measurements extracted from certain regions (e.g., the temporal pole) should be interpreted with caution. Furthermore, possible systematic effects dependent on GBCA dose and delay time after injections should be investigated.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00330-021-08405-8.

Acknowledgements The authors thank the OFAMS study group and the patients who participated in the study. The authors also thank Joost P.A. Kuijer PhD. from the Department of Radiology and Nuclear Medicine at Amsterdam UMC, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, location VUmc, for help with retrieving and interpreting information regarding MRI scanner details.

Funding Open access funding provided by University of Bergen (incl Haukeland University Hospital). K. Wesnes has received funding from Western Norway Regional Health Authority (grant no. #912020).

The MS Center Amsterdam is funded through a program grant of the Dutch MS Research Foundation (grant no. MS 18-358f).

D.R. van Nederpelt is funded by a joint grant from the Dutch MS Research Foundation and ZonMW (grant no. 446002506) and a grant from Health Holland Life Sciences & Health (grant no. LSHM19053).

I. Brouwer has been supported by the Dutch MS Research Foundation.

Ø. Torkildsen and K.M. Myhr are funded by Neuro-SysMed at Haukeland University Hospital and the University of Bergen by a grant from the Research Council of Norway, project number 288164.

F. Barkhof is supported by the National Institute for Health Research (NIHR) biomedical research centre at University College London Hospitals NHS Foundation Trust (UCLH).

Declarations

Guarantor The scientific guarantor of this publication is Hugo Vrenken, PhD.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

I.A. Lie declares no disclosures relevant to the manuscript.

E. Kerklingh declares no disclosures relevant to the manuscript.

K. Wesnes has received unrestricted research grants from Novartis and Biogen, and speaker honoraria from Biogen.
D.R. van Nederpelt declares no disclosures relevant to the manuscript.
I. Brouwer has received research support from Merck KGaA, Novartis, and Teva.
Ø. Torkildsen has received research grants and speaker honoraria from Biogen, Roche, Novartis, Merck, and Sanofi.
K.M. Myhr has received unrestricted research grants to his institution; scientific advisory board, and speaker honoraria from Almirall, Biogen, Genzyme, Merck, Novartis, Roche, and Teva; and has participated in clinical trials organised by Biogen, Merck, Novartis, and Roche.
F. Barkhof has received compensation for consulting services and/or speaking activities from Bayer, Biogen Idec, Merck Serono, Novartis, Roche, Teva, Bracco, and IXICO.
L. Bø has received unrestricted research grants to his institution and/or scientific advisory board or speaker's honoraria from Almirall, Biogen, Genzyme, Merck, Novartis, Roche, and Teva; and has participated in clinical trials organised by Biogen, Merck, Novartis, Roche, and Genzyme.
H. Vrenken has received research grants from Pfizer, Merck Serono, Novartis, and Teva, speaker honoraria from Novartis, and consulting fees from Merck Serono; all funds were paid directly to his institution.

Statistics and biometry No complex statistical methods were used in this paper.

Informed consent Written informed consent was obtained from all subjects (patients) in this study.

Ethical approval The study was approved by the Regional Committee for Medical and Health Research Ethics in Western Norway Regional Health Authority (clinical trials.gov, identifier: NCT00360906).

Study subjects or cohorts overlap Some study subjects or cohorts have been previously reported in "ω-3 fatty acid treatment in multiple sclerosis (OFAMS study): *a randomised, double-blind, placebo-controlled trial* " by Torkildsen et al., Arch Neurol, 2012. 69(8): p. 1044–51, and in "Low vitamin D, but not tobacco use or high BMI, is associated with long-term disability progression in multiple sclerosis" by Wesnes et al., *Mult Scler Relat Disord*, 2021. 50: p. 102801.

Methodology
- Retrospective
- cross-sectional
- multicentre study

References

1. Popescu V, Klaver R, Voorn P et al (2015) What drives MRI-measured cortical atrophy in multiple sclerosis? Mult Scler 21:1280–1290
2. Horakova D, Dwyer MG, Havrdova E et al (2009) 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
3. Zivadinov R, Bergsland N, Dolezal O et al (2013) Evolution of cortical and thalamus atrophy and disability progression in early relapsing-remitting MS during 5 years. AJNR Am J Neuroradiol 34:1931–1939
4. Fisher E, Lee J-C, Nakamura K, Rudick RA (2008) Gray matter atrophy in multiple sclerosis: a longitudinal study. Ann Neurol 64:255–265
5. Sotirchos ES, Gonzalez-Caldito N, Dewey BE et al (2020) Effect of disease-modifying therapies on subcortical gray matter atrophy in multiple sclerosis. Mult Scler 26:312–321
6. Filippi M, Rocca MA, Pagani E et al (2014) Placebo-controlled trial of oral laquinimod in multiple sclerosis: MRI evidence of an effect on brain tissue damage. J Neurol Neurosurg Psychiatry 85:851–858
7. Gaetano L, Haring DA, Radue EW et al (2018) Fingolimod effect on gray matter, thalamus, and white matter in patients with multiple sclerosis. Neurology 90:e1324–e1332
8. Kuperberg GR, Broome MR, McGuire PK et al (2003) Regionally localized thinning of the cerebral cortex in schizophrenia. Arch Gen Psychiatry 60:878–888
9. Rosas HD, Liu AK, Hersch S et al (2002) Regional and progressive thinning of the cortical ribbon in Huntington's disease. Neurology 58:695–701
10. Tustison NJ, Cook PA, Klein A et al (2014) Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. Neuroimage 99:166–179
11. Schmidt MF, Storrs JM, Freeman KB et al (2018) A comparison of manual tracing and FreeSurfer for estimating hippocampal volume over the adult lifespan. Hum Brain Mapp 39:2500–2513
12. Cardinale F, Chinnici G, Bramerio M et al (2014) Validation of FreeSurfer-estimated brain cortical thickness: comparison with histologic measurements. Neuroinformatics 12:535–542
13. Eshaghi A, Prados F, Brownlee WJ et al (2018) Deep gray matter volume loss drives disability worsening in multiple sclerosis. Ann Neurol 83:210–222
14. Steenwijk MD, Geurts JJ, Daams M et al (2016) Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. Brain 139:115–126
15. Calabrese M, Atzori M, Bernardi V et al (2007) Cortical atrophy is relevant in multiple sclerosis at clinical onset. J Neurol 254:1212–1220
16. Zijdenbos AP, Forghani R, Evans AC (2002) Automatic "pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. IEEE Trans Med Imaging 21:1280–1291
17. Nakamura K, Fox R, Fisher E (2011) CLADA: cortical longitudinal atrophy detection algorithm. Neuroimage 54:278–289
18. Smith SM, Zhang Y, Jenkinson M et al (2002) Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage 17:479–489
19. Rovira À, Wattjes MP, Tintoré M et al (2015) MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis—clinical implementation in the diagnostic process. Nat Rev Neurol 11:471–482
20. Saade C, Bou-Fakhredin R, Yousem DM, Asmar K, Naffaa L, El-Merhi F (2018) Gadolinium and multiple sclerosis: vessels, barriers of the brain, and glymphatics. AJNR Am J Neuroradiol 39:2168–2176
21. Rogosnitzky M, Branch S (2016) Gadolinium-based contrast agent toxicity: a review of known and proposed mechanisms. Biometals 29:365–376
22. Kanda T, Oba H, Toyoda K, Kitajima K, Furui S (2016) Brain gadolinium deposition after administration of gadolinium-based contrast agents. Jpn J Radiol 34:3–9

23. Hannoun S, Baalbaki M, Haddad R et al (2018) Gadolinium effect on thalamus and whole brain tissue segmentation. Neuroradiology 60:1167–1173
24. Warntjes JB, Tisell A, Landtblom AM, Lundberg P (2014) Effects of gadolinium contrast agent administration on automatic brain tissue classification of patients with multiple sclerosis. AJNR Am J Neuroradiol 35:1330–1336
25. Torkildsen O, Wergeland S, Bakke S et al (2012) omega-3 fatty acid treatment in multiple sclerosis (OFAMS Study): a randomised, double-blind, placebo-controlled trial. Arch Neurol 69:1044–1051
26. McDonald WI, Compston A, Edan G et al (2001) Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol 50:121–127
27. Schmidt P, Gaser C, Arsic M et al (2012) An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. Neuroimage 59:3774–3783
28. Battaglini M, Jenkinson M, De Stefano N (2012) Evaluating and reducing the impact of white matter lesions on brain volume measurements. Hum Brain Mapp 33:2062–2071
29. Fischl B (2012) FreeSurfer Neuroimage 62:774–781
30. Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9:179–194
31. Desikan RS, Ségonne F, Fischl B et al (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31:968–980
32. Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ (2017) MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. PLoS One 12:e0184661
33. Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging 20:45–57
34. Giavarina D (2015) Understanding Bland Altman analysis. Biochem Med 25:141–151
35. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15:155–163
36. Neacsu V, Jasperse B, Korteweg T et al (2008) Agreement between different input image types in brain atrophy measurement in multiple sclerosis using SIENAX and SIENA. J Magn Reson Imaging 28:559–565
37. Liu Y, Lukas C, Steenwijk MD et al (2016) 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
38. Han X, Jovicich J, Salat D et al (2006) Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage 32:180–194
39. Fennema-Notestine C, Ozyurt IB, Clark CP et al (2006) Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: effects of diagnosis, bias correction, and slice location. Hum Brain Mapp 27:99–113
40. McCarthy CS, Ramprashad A, Thompson C, Botti J-A, Coman IL, Kates WR (2015) A comparison of FreeSurfer-generated data with and without manual intervention. Front Neurosci. https://doi.org/10.3389/fnins.2015.00379
41. Fischl B, Dale AM (2000) Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci USA 97:11050–11055
42. Oguz I, Cates J, Fletcher T et al (2008) Cortical correspondence using entropy-based particle systems and local features.https://doi.org/10.1109/ISBI.2008.4541327
43. Panman JL, To YY, van der Ende EL et al (2019) Bias introduced by multiple head coils in MRI research: an 8 channel and 32 channel coil comparison. Front Neurosci. https://doi.org/10.3389/fnins.2019.00729
44. Li H, Smith SM, Gruber S et al (2018) Combining multi-site/multi-study MRI data: linked-ICA denoising for removing scanner and site variability from multimodal MRI data. bioRxiv. https://doi.org/10.1101/337576:337576
45. Battaglini M, Gentile G, Luchetti L et al (2019) Lifespan normative data on rates of brain volume changes. Neurobiol Aging 81:30–37
46. Guo C, Ferreira D, Fink K, Westman E, Granberg T (2019) 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
47. Platten M, Martola J, Fink K, Ouellette R, Piehl F, Granberg T (2020) MRI-based manual versus automated corpus callosum volumetric measurements in multiple sclerosis. J Neuroimaging 30:198–204
48. Calabrese M, Rinaldi F, Grossi P, Gallo P (2011) Cortical pathology and cognitive impairment in multiple sclerosis. Expert Rev Neurother 11:425–432
49. Inglese M, Petracca M, Mormina E et al (2017) Cerebellar volume as imaging outcome in progressive multiple sclerosis. PLoS One 12:e0176519
50. Chu R, Kim G, Tauhid S, Khalid F, Healy BC, Bakshi R (2018) Whole brain and deep gray matter atrophy detection over 5 years with 3T MRI in multiple sclerosis using a variety of automated segmentation pipelines. PLoS One 13:e0206939

Supplemental data

Contents

Table e1. MRI measurement values obtained by FreeSurfer v.6.0.1

Table e1. MRI measurement values obtained by FreeSurfer v.6.0.1					
MRI measurement	Mean value pre-contrast (SD)	Mean value post-contrast (SD)	Mean difference [a] (SD)	Percent increase/decrease (SD)	ICC (95% confidence interval)
Total grey matter volume (mL)	602.49 (64.13)	624.40 (65.60)	+21.91 (15.57)***	+3.69 (2.64) %	0.985 (0.965-0.994)
Total white matter volume (mL)	454.28 (61.60)	441.17 (59.54)	-13.11 (7.94)***	-2.86 (1.64) %	0.996 (0.990-0.998)
Total deep grey matter volume (mL)	51.14 (5.86)	55.36 (5.53)	+4.22 (2.68)***	+8.58 (6.14) %	0.942 (0.859-.976)
Left Thalamus volume (mL)	6.44 (0.87)	7.89 (1.02)	+1.44 (0.62)***	+22.90 (11.06) %	0.880 (0.711-.950)
Right Thalamus volume (mL)	6.26 (0.89)	7.33 (0.99)	+1.07 (0.58)***	+17.67 (10.07) %	0.896 (0.749-.957)
Mean cortical thickness left hemisphere (mm)	2.35 (0.17)	2.58 (0.18)	+0.23 (0.04)***	+9.60 (1.91) %	0.984 (0.962-0.993)
Mean cortical thickness right hemisphere (mm)	2.35 (0.17)	2.60 (0.17)	+0.25 (0.06)***	+10.74 (2.73) %	0.971 (0.930-0.988)

Abbreviations: SD = standard deviation, ICC = intra-class correlation coefficient for consistency, mL = millilitre, mm = millimetre

[a] Paired t-test

***p<0.0001

Table e2. MRI measurement values obtained from 1.5 and 3.0 T scanners.

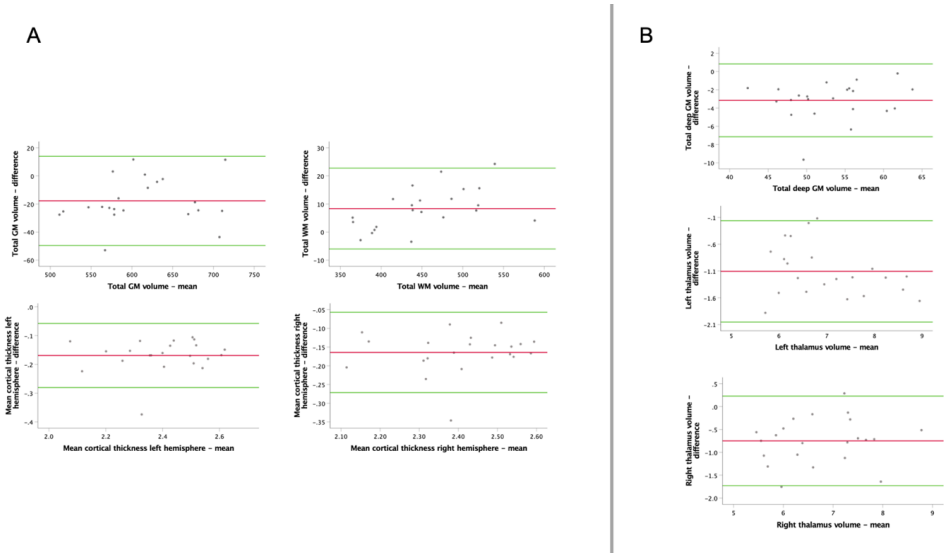
Table e2. MRI measurement values obtained from 1.5 and 3.0 T scanners.						
Field strength	1.5 T (17 patients)			3.0 T (5 patients)		
MRI measure	Mean value pre-contrast (SD)	Mean value post-contrast (SD)	Mean difference[a,b] (SD)	Mean value pre-contrast (SD)	Mean value post-contrast (SD)	Mean difference[a,b] (SD)
Total grey matter volume (mL)	593.44 (63.46)	616.75 (63.76)	23.32 (13.73)**	633.46 (52.95)	632.48 (48.78)	-0.98 (7.35)
Total white matter volume (mL)	459.22 (65.22)	451.68 (62.34)	-7.55 (6.88)**	449.71 (59.39)	438.60 (52.16)	-11.11 (9.04)
Total deep grey matter volume (mL)	51.11 (6.16)	54.05 (6.04)	2.93 (1.50)**	53.14 (5.20)	57.07 (3.26)	3.93 (3.45)
Left Thalamus volume (mL)	6.48 (0.94)	7.49 (1.18)	1.02 (0.49)**	6.47 (1.00)	7.87 (0.72)	1.41 (0.34)*
Right Thalamus volume (mL)	6.27 (1.01)	6.90 (.88)	0.64 (0.45)**	6.72 (0.92)	7.84 (0.70)	1.12 (0.53)*
Mean cortical thickness left hemisphere (mm)	2.28 (0.15)	2.45 (0.15)	0.18 (0.06)**	2.46 (0.04)	2.60 (0.05)	0.14 (0.04)*
Mean cortical thickness right hemisphere (mm)	2.28 (0.14)	2.46 (0.14)	0.17 (0.06)**	2.47 (0.04)	2.60 (0.04)	0.14 (0.03)*

Abbreviations: MRI = magnetic resonance imaging, T = tesla, SD = standard deviation, mL = millilitres, mm = millimetre.

[a] Paired t-test

[b]The results of the performed t-tests for each field strength are viewed as exploratory and should not be directly compared, as patients were only scanned on one scanner. Furthermore, at total of 5 different 1.5T scanners, and 2 different 3T scanners were used.
**p<0.001, *p<0.01

Figure e1. *Bland-Altman plots, illustrating a systematic difference in global (A) and regional (B) MRI measurements, but no proportional bias.*



eAppendix 1 - MRI protocol

Magnetic field strength: Preferably 3 Tesla (T) if available, alternatively 1,5 T
Standard head coil
Positioning: supine

Protocol for MR cerebrum

(Siemens Prisma 3T indicated in brackets; for other configurations, see ADNI/Helse Vest protocols in links below)

1. Localizer
   (Prisma: 3 slices, gap 20%, sagittal (SAG), field of view (FOV) 300mm, matrix (M)=256x256, echo time (TE)/ repetition time (TR)=4/8.6 ms, flip angle (FA) 20, Avg 2)
2. Alternatively quick T2 transverse (TRA)/SAG recording to localize corpus callosum (parameters of choice)
3. Echo planar (EP) diffusion-weighted imaging (DWI), TRA, 5mm slice/ 20 % gap, resolution: (e.g., 1.5T 1.6x1.6x5 mm3)
   Angled after the anterior and posterior part of corpus callosum/the hard palate. b= 0, 1000. Apparent diffusion coefficient (ADC) map.
   (Prisma: Resolve, TRA, 28sl, gap 30 %, FOV 220mm, TE/TR=54/3700ms, M=160x160, 4mm).
4. T2 3D fluid-attenuated inversion recovery (FLAIR) SAG resolution: 1x1x1mm3. Reconstructed in three planes with 0% gap. Alternative to 3D is 2D recording: 2D coronal (COR) FLAIR and SAG T2 FLAIR
   (Prisma: Space-IR, SAG, 192 sl, FOV = 256mm, TE/TR = 386/5000ms, M=256x256)
5. T1 3D magnetization prepared rapid gradient echo (MPRAGE) SAG resolution: 1x1x1mm3. Reconstructed in three planes with 0 % gap. Recordings are done 5 minutes after intravenous contrast injection. (I possible also the same T1 recorded before intravenous contrast injection for optimal segmentation) (Prisma: MPRAGE, SAG, TE/TR/TI= 2.28/1800/900 ms, M=256x256, FA=8)
6. If available 3T MRI; diffusion tensor imaging (DTI) recording: standard echo-planar-imaging DTI (b=0,1000), 6-64 directions, e.g., 5 b=0 and 25 directions b=1000.

Links for protocols:
Helse Vest RHF (the Western Norway Regional Health Authority): https://helse-vest.no/seksjon/radiologiske-prosedyrar/Documents/MR%20nevro/MR%20ms.pdf
ADNI (Alzheimer's Disease Neuroimaging Initiative):
http://adni.loni.usc.edu/methods/documents/mri-protocols/

eAppendix 2 - Summary of FreeSurfer procedures

FreeSurfer uses a combined volume-based and surface-based approach to automatically segment the brain and to calculate volume and average cortical thickness in defined regions of interest. Included in the preprocessing steps are removal of non-brain tissue, registration of the structural volume with the Talairach atlas, assigning neuroanatomical labels to cortical and subcortical regions [1], intensity normalization, tessellation of the white/grey matter boundary, automated topology correction and surface deformation routines to optimally create white/grey and grey/cerebrospinal fluid surface models. These surface models are inflated and then registered to a spherical atlas, to match individual cortical folding patterns to cortical geometry across subjects, based on gyral and sulcal structure [2]. The closest distance from the white/grey boundary to the grey/cerebrospinal fluid boundary at each surface's vertex is defined as the thickness. Lastly, the final cortical and subcortical GM volumes are automatically assigned neuroanatomical labels based on probabilistic global and local spatial information estimated from a manually labelled training set [1,3].

References

1 Fischl B, Salat DH, Busa E et al (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33:341-355

2 Desikan RS, Ségonne F, Fischl B et al (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31:968-980

3 Fischl B, van der Kouwe A, Destrieux C et al (2004) Automatically parcellating the human cerebral cortex. Cereb Cortex 14:11-22

III

OPEN ACCESS

Original research

Serum neurofilament as a predictor of 10-year grey matter atrophy and clinical disability in multiple sclerosis: a longitudinal study

Ingrid Anne Lie ,[1,2] Sezgi Kaçar ,[3] Kristin Wesnes,[2,4] Iman Brouwer,[3] Silje S Kvistad,[2,5] Stig Wergeland,[2,6] Trygve Holmøy,[7,8] Rune Midgard ,[9] Alla Bru,[10] Astrid Edland,[11] Randi Eikeland,[12,13] Sonia Gosal,[14] Hanne F Harbo,[7,15] Grethe Kleveland,[16] Yvonne S Sørenes,[17] Nina Øksendal,[18] Kristin N Varhaug,[1,2] Christian A Vedeler ,[1,2] Frederik Barkhof,[3,19] Charlotte E Teunissen,[20] Lars Bø,[1,21] Øivind Torkildsen,[1,2] Kjell-Morten Myhr,[1,2] Hugo Vrenken[3]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jnnp-2021-328568).

For numbered affiliations see end of article.

Correspondence to
Ms Ingrid Anne Lie, Department of Clinical Medicine, University of Bergen, Bergen, Norway; ingrid.lie@uib.no

Received 3 December 2021
Accepted 18 May 2022
Published Online First 1 June 2022

ABSTRACT

Background The predictive value of serum neurofilament light chain (sNfL) on long-term prognosis in multiple sclerosis (MS) is still unclear.

Objective Investigate the relation between sNfL levels over a 2-year period in patients with relapsing-remitting MS, and clinical disability and grey matter (GM) atrophy after 10 years.

Methods 85 patients, originally enrolled in a multicentre, randomised trial of ω–3 fatty acids, participated in a 10-year follow-up visit. sNfL levels were measured by Simoa quarterly until month 12, and then at month 24. The appearance of new gadolinium-enhancing (Gd+) lesions was assessed monthly between baseline and month 9, and then at months 12 and 24. At the 10-year follow-up visit, brain atrophy measures were obtained using FreeSurfer.

Results Higher mean sNfL levels during early periods of active inflammation (Gd+ lesions present or recently present) predicted lower total (β=−0.399, p=0.040) and deep (β=−0.556, p=0.010) GM volume, lower mean cortical thickness (β=−0.581, p=0.010) and higher T2 lesion count (β=0.498, p=0.018). Of the clinical outcomes, higher inflammatory sNfL levels were associated with higher disability measured by the dominant hand Nine-Hole Peg Test (β=0.593, p=0.004). Mean sNfL levels during periods of remission (no Gd+ lesions present or recently present) did not predict GM atrophy or disability progression.

Conclusion Higher sNfL levels during periods of active inflammation predicted more GM atrophy and specific aspects of clinical disability 10 years later. The findings suggest that subsequent long-term GM atrophy is mainly due to neuroaxonal degradation within new lesions.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ There is increasing evidence to support the use of serum neurofilament light chain (sNfL), as a marker of acute inflammatory axonal damage, to monitor short-term disease activity, treatment response and disability progression in multiple sclerosis (MS). However, whether sNfL levels also predict disease progression and neurodegeneration over several years, and even decades, is less clear.

WHAT THIS STUDY ADDS

⇒ We found that higher sNfL levels measured during periods of active inflammation predicted lower total grey matter (GM) volume, deep GM volume and cortical thickness and higher T2 lesion count after 10 years in patients with relapsing-remitting MS (RRMS). Higher sNfL levels were also associated with higher disability measured by the dominant hand Nine-Hole Peg Test.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE AND/OR POLICY

⇒ As long-term atrophy progression in patients with RRMS seems to be driven by focal inflammatory damage, measuring sNfL levels during relapses may be a way to quantify the extent of ongoing axonal injury, possibly indicating the risk of future disease progression. This added information may support clinicians in subsequent monitoring and treatment decisions.

To cite: Lie IA, Kaçar S, Wesnes K, *et al*. *J Neurol Neurosurg Psychiatry* 2022;**93**:849–857.

INTRODUCTION

The pathological mechanisms in multiple sclerosis (MS) are highly complex, affecting both white matter (WM) and grey matter (GM) structures throughout the central nervous system.[1] Inflammatory and neurodegenerative processes both seem to play a role in disease progression and disability accumulation,[2–4] but there is large variability between patients and disease phenotypes.[4] This pathophysiological and clinical heterogeneity underlines the need for robust biomarkers predicting future clinical disability. At the same time, this heterogeneity poses a challenge in developing such markers, as they should reliably capture and differentiate the various ongoing disease processes.[5]

Neurofilaments are proposed candidate biomarkers, reflecting axonal injury.[6] These proteins are major components of the axonal cytoskeleton

and are released into the extracellular fluid when neuroaxonal damage occurs.[6] The neurofilament protein consists of multiple, differently sized subunits, of which the neurofilament light chain (NfL) assay is the most widely researched.[7] NfL levels can be determined in blood serum or plasma, and serum NfL (sNfL) levels strongly correlate with CSF NfL levels.[8] The suggested dynamic equilibrium between the two body fluids makes NfL a candidate biomarker, because reliable measurements can be obtained by venepuncture, rather than the more invasive lumbar puncture.

Elevated sNfL levels have been shown to reflect acute axonal damage during active inflammation,[9] and increasing evidence support the use of sNfL to monitor short-term disease activity, treatment response and disability progression.[10] Whether sNfL levels also predict disease progression and neurodegeneration over several years, and even decades, is less clear.[6 10–12] Associations between sNfL and long-term disability progression are not consistent,[13 14] and although some studies have found higher sNfL levels to be associated with brain[13 15 16] and GM atrophy,[17–19] studies with extensive follow-up time are few, especially studies considering GM atrophy.[17 19] Clarifying the properties of NfL as a predictor of long-term neurodegeneration is further complicated by the dynamic nature of MS pathophysiological processes: elevated NfL levels during periods with active inflammation mainly reflect the extent of ongoing acute axonal damage, rather than any simultaneous neurodegenerative processes.[20] Furthermore, inflammatory activity and axonal damage persist several months after the appearance of a gadolinium-enhancing (Gd+) lesion, causing a prolonged elevation of the NfL level.[9] If and how this variability affects the relation between NfL levels and long-term future disability and brain atrophy is not clear.[12 17] As one patient with relapsing-remitting MS (RRMS) may experience periods of both remission and active inflammation, attempts to separate and explore the predictive value of sNfL levels during these periods may clarify pathophysiological disease mechanisms, and be of clinical relevance (eg, deciding optimal timepoints for sNfL measurements). By separately analysing sNfL levels obtained during, and outside of episodes of evident inflammatory activity (ie, Gd+ lesions) over a 2-year period, the present study aims to investigate how periods of acute disease activity compare to more silent periods in RRMS in predicting clinical disability and GM atrophy, measured after 10 years.

MATERIALS AND METHODS

Participants

The included patients originally participated in a multicentre trial of ω−3 fatty acids in MS (the OFAMS Study), which has previously been described in detail.[21]

In the trial, 92 patients with RRMS were followed over 24 months, for the first 6 months randomised to either ω−3 fatty acids monotherapy or placebo. Starting at 6 months, both treatment groups received additional treatment with subcutaneously administered interferon beta-1a, 44 µg, three times weekly for the remaining 18 months of the trial. Patients attended regular follow-up visits for biochemical, radiological and clinical examinations, including the Expanded Disability Status Scale (EDSS), timed 25-foot walk test (T25FW), the dominant and non-dominant hand Nine-Hole Peg Test (D9-HPT and ND9-HPT) and the Paced Auditory Serial Addition Test (PASAT). All available patients in the OFAMS Study were invited to a 10-year follow-up visit, of which 85 (92%) accepted.[22] All biochemical, radiological and clinical examinations from the OFAMS Study were repeated at their local study site, with the addition of the oral Symbol Digit Modalities Test (SDMT). Between the OFAMS Study and the 10-year follow-up visit, the participants had received treatment and monitoring as advised by their treating neurologist as part of routine care.

Serum sampling and analysis

Serum samples collected during the OFAMS Study were stored at −80°C. As previously described,[23] sNfL levels were measured in duplicates, from samples collected at baseline (BL) and at months 3, 6, 9, 12 and 24, using a Simoa assay and according to the manufacturer's instruction (Quanterix, Billerica, USA).

MRI data and analysis

The OFAMS Study

During the trial, patients underwent MRI imaging at BL, monthly for the first 9 months, and thereafter at month 12 and 24. MRI was performed at each study site using a 1.5 Tesla (T) MRI scanner with the standard head coil. After intravenous injection of gadolinium-based contrast agent, the imaging protocol included a 2D sagittal fluid-attenuated inversion recovery (FLAIR) (resolution: $0.98\times0.98\times1\,mm^3$, echo time (TE)/repetition time (TR)=100/6000–10000 ms, number of excitations (NEX) 2, slice thickness 4 mm), 2D axial T1-weighted images (resolution: $0.49\times0.49\times1\,mm^3$, TE/TR=10–20/500-750 ms, NEX 2, slice thickness 4 mm) as well as sagittal 3D T1-weighted spoiled gradient echo (Fast Field Echo (FFE)/Fast Low Angle Shot (FLASH)) images (resolution: $0.98\times0.98\times1\,mm^3$, TE/TR=4.6/20 ms, flip angle 25°, NEX 1, slice thickness 1 mm).

Blinded assessment of the T2 and Gd+ lesion count (LC) at BL, and the appearance of new Gd+ lesions was conducted by two experienced neuroradiologists.

The 10-year follow-up visit

Imaging was performed at the different study sites, on a 3T MRI scanner if available, alternatively using a 1.5 T MRI scanner, with a standard head coil. The following MRI sequences were acquired: a T2-weighted 3D sagittal FLAIR (resolution: $1\times1\times1\,mm^3$, TE/TR/inversion time (TI)=386/5000/1.65–2.2 ms) and a postcontrast T1-weighted 3D sagittal magnetization prepared rapid gradient echo sequence (resolution: $1\times1\times1\,mm^3$, TE/TR/TI=2.28/1800/900 ms, flip angle 8°).

Lesion segmentation and morphological reconstruction

A detailed description of these methods has recently been described[24] and is available in online supplemental appendix 1. Briefly, on images obtained at the 10-year follow-up visit, lesion segmentation was done on FLAIR images using Lesion Segmentation Tool (V.2.0.15; http://applied-statistics.de/lst.html),[25] and morphological reconstruction was performed with FreeSurfer (V.7.1.1; http://surfer.nmr.mgh.harvard.edu/) on T1-weighted images.

Calculation of sNfL levels

Mean sNfL levels were calculated, for each patient, for three different settings: 'overall mean sNfL level', from all samples collected between BL and month 24; 'mean inflammatory sNfL level', from samples collected within 2 months after the presence of a Gd+ lesion, or less than 2 weeks before the appearance of a Gd+ lesion (if collected more than 1 week after last MRI scan); and 'mean non-inflammatory sNfL level', from samples collected more than 2 months after the appearance of a Gd+ lesion and more than 2 weeks before the appearance of a new

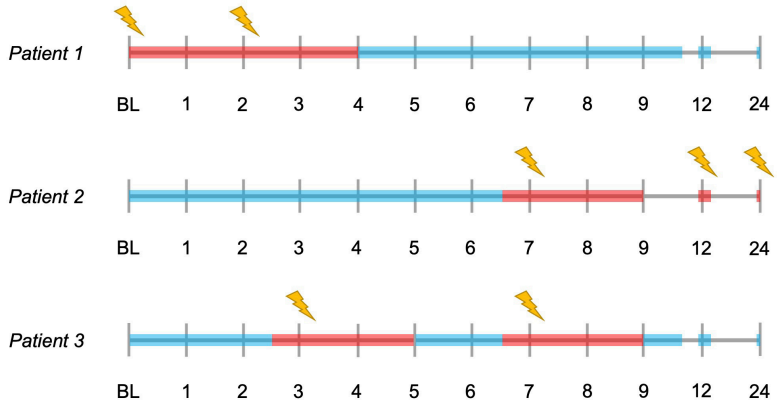


Figure 1 Illustrated examples of time periods where the collected serum neurofilament light chain (sNfL) levels are defined as 'inflammatory' or 'non-inflammatory'. The timelines represent the MRI visits during the OFAMS Study; visits with a new gadolinium-enhancing lesion are marked with a lightning symbol. sNfL levels collected during periods marked in red are defined as inflammatory and levels collected during periods marked in blue are defined as non-inflammatory. With sNfL levels collected approximately at baseline, month 3, 6, 9, 12 and 24; patient 1 has two inflammatory (included in the analysis requiring at least two measurements, excluded from the analysis requiring at least three measurements) and four non-inflammatory sNfL levels (included in both analyses); patient 2 has three inflammatory (included in both analyses) and three non-inflammatory sNfL levels (included in both analyses); and patient 3 has two inflammatory (included in one analysis) and four non-inflammatory (included in both analyses) sNfL levels.

Gd+ lesion (if collected more than 1 week after last MRI scan). Examples of sNfL measurements defined as inflammatory and non-inflammatory are visualised in figure 1. In each patient, the mean inflammatory and non-inflammatory sNfL level was calculated separately for (1) at least two and (2) at least three measurements, when available. Measurements defined as inflammatory or non-inflammatory did not have to be collected at consecutive timepoints. The findings presented here were obtained using the mean of at least three measurements, highly comparable findings using the mean of at least two measurements are presented in the online supplemental tables 1 and 2.

Statistical analysis

Statistical analyses were performed using R software (V.4.0.5). Thalamus volume and mean cortical thickness in the left and right hemisphere were averaged.

To correct for the different study sites and scanner variability, the relationship between overall mean sNfL level and clinical and MRI atrophy measures was investigated by a linear multilevel regression model, corrected for age, sex, disease modifying therapy (DMT) use, estimated total intracranial volume (eTIV) (eTIV only included in analyses regarding MRI volume measures), fraction of MRI scans with new Gd+ lesions (fGd+), BL T2 and Gd+ LC, with study site entered as a random effect.

Between the OFAMS Study and the 10-year follow-up visit, patients underwent therapeutic interventions that varied both between and within patients, in potency, duration and time. A nominal variable was created based on the category (similar to those proposed in a recent study[26]) of DMT(s) used during the follow-up: (1) only used platform compounds (interferon beta and glatiramer acetate preparations), (2) ever used oral therapies (teriflunomide, dimethyl fumarate, fingolimod) and (3) ever used high efficiency monoclonal antibody therapies, chemotherapies or haematopoietic stem cell therapy.

For the relation between mean inflammatory and non-inflammatory sNfL levels and clinical and MRI atrophy measures, linear regression models were used, as entering the study site as a random effect did not improve the model. The first model (model 1) included mean inflammatory sNfL level, fGd+, age, sex, DMT use, eTIV, BL T2 and Gd+ LC as independent variables; the second model (model 2) included non-inflammatory sNfL level, age, sex, DMT use, eTIV, BL T2 and Gd+ LC. Lastly, a modified version of model 1 was used in two exploratory analyses: the first with the mean cortical thickness in the precentral gyrus as the dependent variable, and the second including MRI atrophy measures obtained at month 24 (available in a subset of patients) as a covariate. All independent variables were first entered as covariates and removed by backward elimination if not significant to the model. In case of missing observations, patients were excluded from the respective analyses. Assumptions for linear regression were checked for each final model; if the assumptions were not satisfied, log-linear transformation was performed (eg, logT25FW). The outcome measure EDSS≥4 was investigated by logistic regression. Lastly, the Benjamini-Hochberg method[27] was used to control the false discovery rate (FDR) for multiple hypothesis testing. FDR controlling was performed for the main predictors (overall sNfL, inflammatory sNfL, non-inflammatory sNfL and fGd+) separately, including analyses with both MRI and clinical outcome measures.

RESULTS

Patient characteristics

Of the 85 patients who participated in the 10-year follow-up visit, 78 had serum samples available for sNfL measurement and were included in this study. The mean follow-up time from BL to the 10-year follow-up visit was 12.0 years (±0.6). Table 1 summarises clinical and MRI characteristics of the included patients.

Overall mean sNfL level

Overall mean sNfL level did not predict any long-term MRI or clinical outcome measures, or change in clinical measures from month 24 to the 10-year follow-up (table 2).

Table 1 Demographic, clinical and radiological characteristics

	N	Baseline	Month 24	10-year follow-up visit
Age in years, mean (SD)/median (range)	78			50.05 (8.4)/50.0 (31–70)
Sex, female, N (%)	78	51 (65.4%)		
Time since diagnosis, mean in years (SD)/median (range)	78			14.6 (3.4)/13.7 (11.0–26.1)
Disease phenotype (N)	78	RRMS (78)	RRMS (78)	RRMS (71), SPMS (7)
Type of DMT used during follow-up (N)	78	Only platform compounds* (23), ever used oral therapies† (32), ever used high efficiency monoclonal antibody therapies, chemotherapies, or HSCT‡ (23).		
Study site (number of patients)	78	Site 1 (3), site 2 (16), site 3 (3), site 4 (2), site 5 (1), site 6 (5), site 7 (8), site 8 (13), site 9 (3), site 10 (6), site 11 (2), site 12 (12), site 13 (4).		
EDSS, mean (SD)/median (range)	78/76/77	1.9 (0.8)/2.0 (0.0–4.0)	2.1 (1.2)/2.0 (0.0–5.0)	2.8 (1.6)/2.5 (0.0–8.5)
Mean sNfL level§ (pg/mL), mean (SD)	78	34.8 (14.3)		
Mean inflammatory sNfL level§ (pg/ml), mean (SD)	32	45.5 (21.3)		
Mean non-inflammatory sNfL level§ (pg/mL), mean (SD)	40	30.2 (9.5)		
fGd+, mean (SD)	78	0.32 (0.26)		
Number of MRI scans with new Gd-enhancing lesions, mean (SD)/median (range)	78	3.7 (3.1)/3.0 (0–11)		
Total GM volume (mm^3), mean (SD)	65			630 134.461 (52 453.119)
Total WM volume (mm^3), mean (SD)	65			448 155.938 (50 676.88)
Total deep GM volume (mm^3), mean (SD)	65			55 726.031 (5291.634)
Thalamus volume (mm^3), mean (SD)	65			7786.642 (982.467)
Mean Cth (mm), mean (SD)	65			2.538 (0.128)

*Interferon beta and glatiramer acetate preparations.
†Dimethyl fumarate, teriflunomide, fingolimod.
‡Natalizumab, rituximab, alemtuzumab, mitoxantrone, haematopoietic stem cell therapy.
§Mean sNfL levels measured from serum samples collected from baseline to month 24.
Cth, cortical thickness; DMT, disease modifying therapy; EDSS, Expanded Disability Status Scale; fGd+, fraction of MRI scans with new Gadolinium-enhancing lesion; Gd, Gadolinium; GM, grey matter; HSCT, haematopoietic stem cell therapy; RRMS, relapsing-remitting multiple sclerosis; sNfL, serum neurofilament light; SPMS, secondary progressive multiple sclerosis; WM, white matter.

Table 2 The association of overall mean sNfL level with MRI atrophy and clinical measures at the 10-year follow-up, with a random intercept for study site, corrected for age, sex, DMT use, eTIV, BL T2LC, BL Gd+ LC and fGd+

MRI/clinical measurement	N	B	Std. B	95% CI	P value*	Marginal R^2	Conditional R^2
Total GM volume	65	−471.6	−0.147	−1236.446 to 293.239	0.514	0.385	0.607
Total WM volume	65	−110.9	−0.030	−945.240 to 723.354	0.920	0.380	0.380
Total deep GM volume	65	−78.12	−0.221	−162.299 to 6.054	0.429	0.423	0.513
Thalamus volume	65	−12.778	−0.203	−29.365 to 3.808	0.487	0.276	0.501
Mean Cth	65	−0.002	−0.255	-0.004 to 1.069×10^{-4}	0.782	0.308	0.584
logLesion volume†	68	-2.830×10^{-4}	−0.001	−0.006 to 0.006	0.989	0.351	0.499
Lesion count	68	0.112	0.086	−0.046 to 0.270	0.488	0.272	0.430
EDSS≥4‡	77	0.000	1.000	0.952 to 1.052	0.985		
logT25FW†	72	−0.001	−0.158	−0.004 to 0.001	0.470	0.096	0.373
logChange in T25FW†	70	−0.001	−0.129	-0.003 to 3.932×10^{-4}	0.581	0.062	0.258
logD9-HPT†	71	0.002	0.263	-4.058×10^{-6} to 0.004	1.000	0.309	0.348
logChange in D9-HPT†	69	−0.001	−0.076	−0.005 to 0.003	0.735	0.195	0.229
logND9-HPT†	70	−0.001	−0.073	−0.002 to 0.001	0.670	0.278	0.368
logChange in ND9-HPT†	68	−0.004	−0.234	-0.009 to 3.350×10^{-4}	0.550	0.170	0.239
PASAT	72	0.088	0.112	−0.085 to 0.260	0.550	0.189	0.247
Change in PASAT	70	0.038	0.063	−0.082 to 0.157	0.738	0.143	0.431
Oral SDMT	67	0.110	0.128	−0.094 to 0.314	0.563	0.222	0.434

Marginal R^2: variance explained by fixed effects.
Conditional R^2: variance explained by both fixed and random effects.
*Adjusted p values after controlling the false discovery rate (FDR) for multiple hypothesis testing.
†Dependent variable log transformed due to non-normality (log-linear transformation).
‡Analysed by logistic regression, regression coefficient (B), odds ratio (Std. B) and 95% CI of odds ratio reported.
B, beta; BL, baseline; Cth, cortical thickness; D9-HPT, dominant hand Nine-Hole Peg Test; DMT, disease modifying therapy; EDSS, Expanded Disability Status Scale; eTIV, estimated total intracranial volume; fGd+, fraction of MRI scans with new Gadolinium-enhancing lesion; Gd+, gadolinium-enhancing; GM, grey matter; LC, lesion count; ND9-HPT, non-dominant hand Nine-Hole Peg Test; PASAT, Paced Auditory Serial Addition Test; SDMT, Symbol Digit Modalities Test; sNfL, serum neurofilament light chain; Std, standardised; T25FW, timed 25-foot walk; WM, white matter.

Table 3 Model 1: The association of inflammatory sNfL level and fGd+ with MRI atrophy and clinical measures at the 10-year follow-up, corrected for age, sex, DMT use, eTIV, BL T2LC and Gd+ LC*

		Mean inflammatory sNfL level				fGd+				Full model	
MRI/clinical measure	N	B	Std. B	95% CI	P value†	B	Std. B	95% CI	P value†	R^2 adj.	P value
Total GM volume	25	−850.8	−0.399	−1580.218 to −121.416	**0.040**	91 552.9	0.362	3400.111 to 179 705.771	0.065	0.504	<0.001
Total WM volume	25				NS				NS		
Total deep GM volume	25	−140.31	−0.556	−228.417 to −52.198	**0.010**				NS	0.341	0.004
Thalamus volume	25				NS				NS		
Mean Cth	25	−0.003	−0.581	−0.005 to −0.001	**0.010**				NS	0.308	0.002
logLesion volume‡	28				NS				NS		
logLesion count‡	28	0.004	0.498	0.001 to 0.007	**0.018**				NS	0.220	0.007
EDSS≥4§	31				NS				NS		
logT25FW‡	30				NS				NS		
logChange in T25FW‡	30				NS				NS		
logD9-HPT‡	29	0.004	0.593	0.002 to 0.006	**0.004**				NS	0.411	0.001
logChange in D9-HPT‡	29				NS				NS		
logND9-HPT‡	29				NS				NS		
logChange in ND9-HPT‡	29	−0.006	−0.498	−0.010 to −0.001	**0.024**				NS	0.399	0.002
PASAT	28				NS				NS		
Change in PASAT	28				NS				NS		
Oral SDMT	28				NS				NS		

*Non-significant covariates removed from final model by backward elimination.
†Adjusted p values after controlling the false discovery rate (FDR) for multiple hypothesis testing.
‡Dependent variable log transformed due to non-normality (log-linear transformation).
§Analysed by logistic regression.
adj, adjusted; B, beta; BL, baseline; Cth, cortical thickness; D9-HPT, dominant hand Nine-Hole Peg Test; DMT, disease modifying therapy; EDSS, Expanded Disability Status Scale; eTIV, estimated total intracranial volume; fGd+, fraction of MRI scans with new Gadolinium-enhancing lesion; Gd+, gadolinium-enhancing; GM, grey matter; LC, lesion count; ND9-HPT, non-dominant hand Nine-Hole Peg Test; PASAT, Paced Auditory Serial Addition Test; SDMT, Symbol Digit Modalities Test; sNfL, serum neurofilament light chain; Std, standardised; T25FW, timed 25-foot walk; WM, white matter.

Mean inflammatory sNfL level

The results of the linear regression model including inflammatory sNfL and fGd+ as predictor variables (model 1) are shown in table 3.

Higher mean inflammatory sNfL level predicted lower total GM (standardised β=−0.399, p=0.040) and deep GM (standardised β=−0.556, p=0.010) volume, lower mean cortical thickness (standardised β=−0.581, p=0.010) and higher logT2LC (standardised β=0.498, p=0.018) (figure 2). Of all the clinical outcomes, higher mean inflammatory sNfL level was associated with a higher score (higher disability) on the logD9-HPT (standardised β=0.593, p=0.004) and a lower increase (less disability accumulation) in the logND9-HPT score (standardised β=−0.498, p=0.024) between month 24 and the 10-year follow-up.

Fraction of active MRI scans was not a significant predictor in any of the models (table 3).

Exploratory analyses

In a subset of patients, inflammatory sNfL levels were not associated with any MRI measurement obtained at the 10-year follow-up, after correcting for MRI atrophy measurements obtained at month 24 (online supplemental table 3).

Higher mean inflammatory sNfL levels and D9-HPT scores, but not ND9-HPT scores, were significantly associated with lower cortical thickness in the left and right precentral gyrus (online supplemental table 4).

Mean non-inflammatory sNfL level

The effect of mean non-inflammatory sNfL level on MRI and clinical measures at the 10-year follow-up is shown in table 4. The mean non-inflammatory sNfL level was not associated with any of the MRI measures. For the clinical measures, higher levels were solely associated with a higher SDMT score (better attention score) at the 10-year follow-up (standardised β=0.473, p=0.003).

DISCUSSION

We found that higher mean sNfL level, measured over a 2-year period in patients with RRMS, was not associated with MRI or clinical measures after 10 years. However, when separately assessing mean sNfL levels measured during periods of active inflammation, higher levels associated significantly with lower total GM and deep GM volume, lower cortical thickness, higher T2 LC and higher disability measured by the D9-HPT. Lastly, sNfL levels during remission were not associated with long-term atrophy or disability progression. These findings suggest that sNfL levels during active inflammation may better predict atrophy and disability progression than overall mean sNfL and sNfL levels during remission.

Inflammatory sNfL levels were analysed in samples collected during periods with focal active inflammation, reflecting the extent of acute axonal damage.[9] The association with GM atrophy measured after 10 years, implies that the delayed neurodegeneration in certain GM regions is at least partly secondary to focal inflammatory damage, most likely through anterograde or retrograde neuroaxonal degeneration along WM tracts.[28] An alternative hypothesis could be that the association is based on pseudoatrophy following resolved inflammatory activity, but as pseudoatrophy is shown to mainly affect the WM,[29] this seems less plausible. Elevated NfL levels predicting secondary neurodegeneration have been suggested in previous works,[17 30] finding an association between higher sNfL levels and atrophy progression in deep GM over a 5-year[17] and 6-year[19] follow-up period. Our study supports this further, by assessing inflammatory sNfL levels separately and finding that the associated GM atrophy

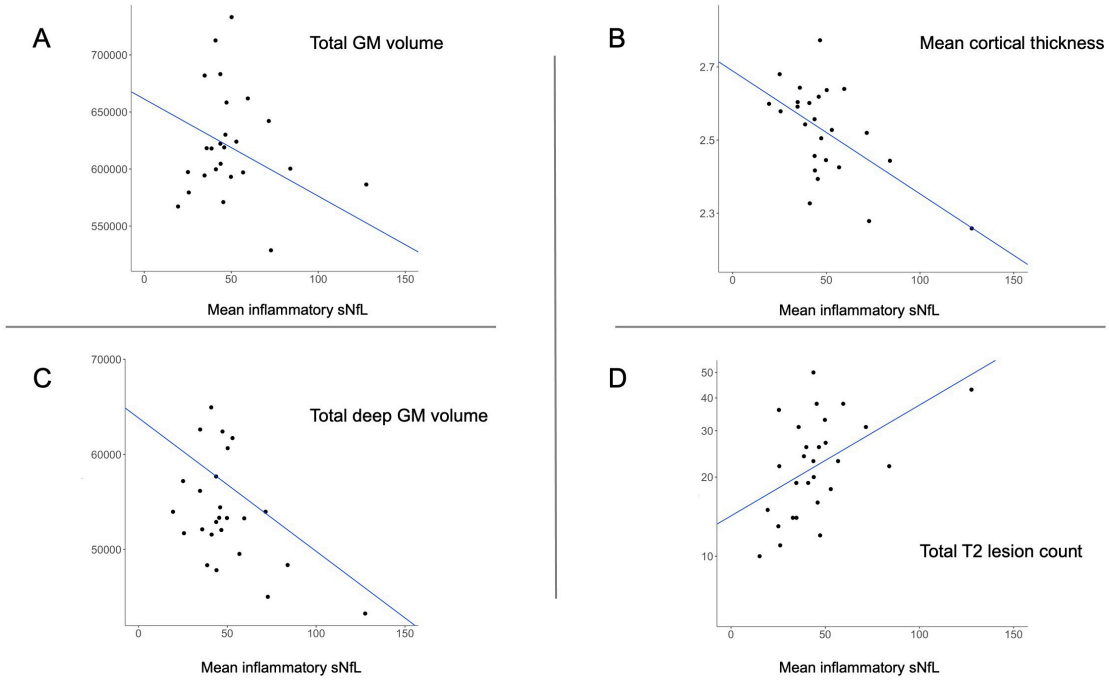


Figure 2 Scatterplots illustrating significant associations between mean inflammatory sNfL level (pg/mL) and (A) total GM volume (mm^3), (B) mean cortical thickness (mm), (C) total deep GM volume (mm^3) and (D) total T2 lesion count (N). The Y-axis is transformed to logarithmic scale to illustrate the absolute lesion count. Cth, cortical thickness; GM, grey matter; sNfL, serum neurofilament light chain.

was located in deep GM and the cerebral cortex, areas known to be highly interconnected through various WM circuits,[31 32] therefore susceptible to degradation.[33 34] When correcting for MRI atrophy measures at month 24, associations with inflammatory sNfL levels were no longer significant. This may imply that the atrophy progression develops relatively early in the disease course. However, this exploratory analysis was conducted in only a subset of patients, and should in future works be repeated in larger cohorts.

We found that higher sNfL levels were associated with a higher score (higher disability) on the logD9-HPT. This result is partly in line with a previous study, showing that the patient group with prominent spinothalamic atrophy progression had higher sNfL levels and developed motor disability faster than the groups with atrophy progression in other regions.[19] The difference between groups was most evident when assessing walking speed (T25FW) and finger dexterity (9-HPT).[19] Hypothetically, the associated disability progression may result from acute disruption of crucial WM tracts (ie, the corticospinal tract) and secondary upstream neurodegeneration in connected GM areas (ie, the primary motor cortex).[34] In our study, higher inflammatory sNfL levels were also associated with a lower change (less disability accumulation) in the logND9-HPT. While this finding does not coincide with the suggested hypothesis, the analysis may have been influenced by statistical power-issues and outliers. Furthermore, in an exploratory analysis (online supplemental table 4), lower cortical thickness in the precentral gyri was associated with both higher inflammatory sNfL levels and higher disability measured by the D9-HPT, but not by the ND9-HPT.

We found no associations between sNfL levels and EDSS. In previous research, the relation between sNfL and EDSS progression over 10 years or more is variable,[13 14 16] suggested to be influenced by the difference in disease severity between cohorts.[10] Our study of a limited number of patients, with relatively low overall disability progression (namely: up to EDSS 2.8), may be affected by the known low sensitivity to change in EDSS,[35] especially for lower scores.

A higher fraction of MRI scans with new Gd+ lesions was not a significant predictor in any of the models. Compared with the results seen for inflammatory sNfL, this lack of significant associations may be due to the less sensitive fractional measure used, based on dichotomised values. Nevertheless, the discrepant results for the two predictors may also mean that future neurodegeneration and disability depend on the extent of axonal damage and location of an episode with a new Gd+ lesion(s), more than the frequency of such episodes.

Except for a positive relationship between non-inflammatory sNfL level and oral SDMT, none of the models for this predictor were significant. This may be influenced by statistical power issues and outliers, as the sample size was small (40 patients), with small overall variability in sNfL levels. As only patients with at least three samplings of non-inflammatory sNfL levels available were included, analyses may also be subject to selection bias, selecting patients with an overall less active disease course (none of the patients had at least three non-inflammatory and inflammatory measurements available). However, repeating the analyses including patients with a minimum of two non-inflammatory sNfL levels, and subsequently patients with periods

Table 4 Model 2: The association of mean non-inflammatory sNfL level with MRI atrophy and clinical measures at the 10-year follow-up, corrected for age, sex, DMT use, eTIV, BL T2LC and BL Gd+ LC*

		Mean non-inflammatory sNfL level				Full model	
MRI/clinical measure	N	B	Std. B	95% CI	P value†	R^2 adj.	P value
Total GM volume	36				NS		
Total WM volume	36				NS		
Total deep GM volume	36				NS		
Thalamus volume	36				NS		
Mean Cth	36				NS		
logLesion volume‡	36				NS		
Lesion count	36				NS		
EDSS≥4§	40				NS		
logT25FW‡	38				NS		
logChange in T25FW‡	38				NS		
logD9-HPT‡	38				NS		
logChange in D9-HPT‡	38				NS		
logND9-HPT‡	37				NS		
logChange in ND9-HPT‡	37				NS		
PASAT	40				NS		
Change in PASAT	40				NS		
Oral SDMT	35	0.548	0.473	0.196 to 0.900	**0.003**	0.380	<0.001

*Non-significant covariates removed from final model by backward elimination.
†Adjusted p values after controlling the false discovery rate (FDR) for multiple hypothesis testing.
‡Dependent variable log transformed due to non-normality (log-linear transformation).
§Analysed by logistic regression.
adj, adjusted; B, beta; BL, baseline; Cth, cortical thickness; D9-HPT, dominant hand Nine-Hole Peg Test; DMT, disease modifying therapy; EDSS, Expanded Disability Status Scale; eTIV, estimated total intracranial volume; Gd+, gadolinium-enhancing; GM, grey matter; LC, lesion count; ND9-HPT, non-dominant hand Nine-Hole Peg Test; PASAT, Paced Auditory Serial Addition Test; SDMT, Symbol Digit Modalities Test; sNfL, serum neurofilament light chain; Std, standardised; T25FW, timed 25-foot walk; WM, white matter.

of both remission and active disease (35 patients), yielded the same results. The results are also in line with a recent study finding no association between NfL level and disease progression in natalizumab-treated patients, after correcting for MRI activity.[12] From these findings, the authors hypothesised that the sensitivity of NfL is too low to capture more subtle neuroaxonal damage not associated with active inflammation.

The findings in our study may have clinical relevance. Long-term outcomes were independently predicted by sNfL levels during inflammatory episodes, and not by the frequency of such episodes during the first 2 years. Hence, measuring sNfL levels during relapses may be a way to quantify the extent of ongoing axonal damage, possibly indicating the risk of permanent disability, either caused by direct axonal damage during active inflammation, or by the delayed secondary neurodegenerative process affecting GM in connected regions. This added information may support clinicians in subsequent monitoring and treatment decisions. Furthermore, the addition of sNfL to treatment response scoring tools[36 37] could possibly increase their predictive value, and should be assessed in future studies.

Correcting for DMT use did not change the associations between sNfL levels and long-term outcomes. However, use of high efficiency therapies (indicating disease activity) over the follow-up was independently associated with disability accumulation measured by the 9-HPT (results not shown). As patients were treated similarly until the conclusion of the OFAMS Study (first treatment naïve, then treated with interferons), this suggests that potent treatment during the first years after diagnosis is important for long-term prognosis, especially in patients with high disease activity.

This study has limitations, the main challenges and suggestions for future research are summarised in table 5. There is a degree of uncertainty in defining sNfL levels as 'inflammatory' or 'non-inflammatory'. Regarding different lesion types, we focused on their relation with Gd+ lesions, as these are strongly associated with active inflammation and NfL release,[10] and can be temporally identified with great certainty. However, at BL, month 12 and 24, there was no MRI scan available from the previous months to decide on recent inflammatory activity, and spinal lesions were not accounted for. After a Gd+ lesion, increased

Table 5 Current research challenges and suggestions for future research

Research challenges	Suggestions for future research
Clarify the temporal relation between sNfL levels and new, enlarging and diminishing lesions, for example, ► T2 hyperintense lesions ► T1 hypointense lesions ► T1 Gd+ hyperintense lesions ► GM lesions ► Spinal lesions	► Prospective studies ► Sufficient sample size ► Extensive follow-up time ► Frequent follow-up visits, including: Imaging techniques suited for analyses of longitudinal lesion and atrophy progression. ► Statistical analyses correcting for known risk factors and modulators of disease progression: Baseline and on-study lesion activity. Previous and on-study therapeutic interventions. Genetic and environmental risk factors. Comorbid conditions. ► Consider using z scores for sNfL derived from a healthy control group or a reference database.[26].
Clarify the temporal relation between sNfL levels and GM atrophy progression, for example, ► Global brain GM atrophy ► Regional brain GM atrophy ► Spinal atrophy	
Clarify the value of sNfL as an independent predictive biomarker of long-term prognosis.	
Establishing sNfL reference values.	

Gd+, gadolinium-enhancing; GM, grey matter; sNfL, serum neurofilament light chain.

sNfL levels may persist for up to 90 days,[9] and a previous study on this patient cohort found elevated sNfL levels up to 1 month before and 2 months after the appearance of Gd+ lesions,[23] indicating that the windows for defining a sNfL measurement as inflammatory or non-inflammatory in the current study may be too narrow and too wide, respectively. Non-inflammatory measurements are at highest risk of misclassification, ideally collected with a wider interval between new lesions, to ensure the levels are not influenced by inflammatory damage. These considerations underline the need to clarify the relationship between the temporal dynamics of NfL levels and the evolution of lesions. With the available data in this cohort, our definitions were set to maximise the contrast between inflammatory and non-inflammatory periods, while still maintaining an acceptable group size. Despite these uncertainties, the associations with long-term outcomes found in this study were clearly different between the two measurements, substantiating the sensitivity of the set definitions. Moreover, the patterns of significant associations were similar when analysing mean inflammatory and non-inflammatory sNfL levels calculated from only two or more measurements, also including patients (35 patients) with both inflammatory and non-inflammatory sNfL levels during the 2-year follow-up.

GM volumes were measured cross-sectionally from data collected at the 10-year follow-up visit and month 24, limiting our ability to conclude on longitudinal atrophy progression. When correcting for atrophy measures obtained at month 24, the associations with GM atrophy after 10 years were no longer significant. This analysis may have been underpowered due to the small sample size, so further investigations in larger patient populations, with regular and more frequent follow-up visits, may clarify the temporal relation between inflammatory WM damage, sNfL levels and GM atrophy. Additionally, future studies should consider the effect of lesion volume and lesion volume change, preferably over longer time periods. In this study, we corrected for Gd+ and T2 LC at BL, as we deemed BL volume measures too unreliable to include, due to the quality of the MRI data (eg, partial brain coverage, large slice thickness, 2D images).

Lastly, atrophy measurements were obtained from postcontrast images, which is not the standard approach for FreeSurfer. However, recent work has shown excellent consistency between values obtained from precontrast and postcontrast images.[24]

Conclusion

Higher sNfL levels during early periods of active inflammation, but probably not during remission, in patients with RRMS predicted GM atrophy and specific aspects of clinical disability 10 years later. The findings suggest that subsequent long-term GM atrophy is mainly due to neuroaxonal degradation induced by acute inflammation.

Author affiliations
[1]Department of Clinical Medicine, University of Bergen, Bergen, Norway
[2]Neuro-SysMed, Department of Neurology, Haukeland University Hospital, Bergen, Norway
[3]Department of Radiology and Nuclear Medicine, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, Amsterdam, The Netherlands
[4]Department of Neurology, St. Olav's University Hospital, Trondheim, Norway
[5]Department of Immunology and Transfusion Medicine, Haukeland University Hospital, Bergen, Norway
[6]Norwegian Multiple Sclerosis Registry and Biobank, Department of Neurology, Haukeland University Hospital, Bergen, Norway
[7]Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[8]Department of Neurology, Akershus University Hospital, Lorenskog, Norway
[9]Department of Neurology, Molde Hospital, Molde, Norway
[10]Department of Neurology, Stavanger University Hospital, Stavanger, Norway
[11]Department of Neurology, Vestre Viken Hospital Trust, Drammen, Norway
[12]Department of Research and Education, Sørlandet Hospital Trust, Kristiansand, Norway
[13]Faculty of Health and Sport Science, University of Agder, Grimstad, Norway
[14]Department of Neurology, Østfold Hospital Kalnes, Grålum, Norway
[15]Department of Neurology, Oslo University Hospital, Oslo, Norway
[16]Department of Neurology, Innlandet Hospital Trust, Lillehammer, Norway
[17]Department of Neurology, Haugesund Hospital, Haugesund, Norway
[18]Department of Neurology, Nordland Hospital Trust, Bodø, Norway
[19]Institutes of Neurology and Healthcare Engineering, UCL, London, UK
[20]Neurochemistry Laboratory, Clinical Chemistry Department, Amsterdam Neuroscience, Amsterdam University Medical Centres, Amsterdam, The Netherlands
[21]Norwegian Multiple Sclerosis Competence Centre, Department of Neurology, Haukeland University Hospital, Bergen, Norway

Acknowledgements The authors thank the OFAMS Study group and the patients who participated in the study.

Contributors Conception or design of the work: IAL, CET and HV. Data acquisition: IAL, KW, IB, SSK, SW, TH, RM, AB, RE, SG, HFH, GK, YSS, NØ, KNV, CAV, ØT, KMM and HV. Analysis and interpretation of data: IAL, SK, FB, CET, LB, ØTT, KMM and HV. Drafting the work: IAL and HV. Revising the work for valuable intellectual content and final approval of the version: IAL, SK, KW, IB, SSK, SW, TH, RM, AB, RE, SG, HFH, GK, YSS, NØ, KNV, CAV, FB, CET, LB, ØT, K-MM and HV. Authors responsible for the overall content as guarantors: IAL and HV.

Funding This study received no specific funding. The OFAMS baseline study has been funded by Pronova Biocare; Amersham Health, Norway; Merck Serono, Norway; the Western Norway Regional Health Authority; and Norwegian Multiple Sclerosis Society (no award/grant number). The OFAMS follow-up study has received unrestricted research grants from Novartis and The Independent order of Odd Fellows; has been funded by the Western Norway Regional Health Authority (grant number 912020); and has been financial supported by Neuro-SysMed (Center of excellence for clinical treatment research) hosted by Haukeland University Hospital and funded by grants from the Research Council of Norway (grant number 288164). The MS Center Amsterdam is funded through a program grant of the Dutch MS Research Foundation (grant no. MS 18-358f). IB has been supported by the Dutch MS Research Foundation (no award/grant number). FB is supported by the National Institute for Health Research (NIHR) biomedical research centre at University College London Hospitals NHS Foundation Trust (UCLH) (no award/grant number). ØT and K-MM are funded by Neuro-SysMed at Haukeland University Hospital and University of Bergen by grant from the Research Council of Norway (grant number 288164).

Competing interests The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article. AB, SG, GK, YSS, KNV and CAV declares no disclosures relevant to the manuscript. IAL has received research grants from the Meltzer Research Fund, Gerda Meyer Nyquist Guldbrandson & Gert Meyer Nyquists Legat and the Independent Order of Odd Fellows. KW has received unrestricted research grants from Novartis and Biogen, research grant from the Independent Order of Odd Fellows and speaker honoraria from Biogen. IB has received research support from Merck KGaA, Novartis, and Teva. SW has received speaker honoraria from and served on scientific advisory boards for Biogen, Janssen-Cilag, Sanofi and Novartis. SSK received unrestricted research grants from Novartis, Biogen. TH has received speaker honoraria, research support/grants and participated in clinical trials for Biogen, Merck, Sanofi, Bristol Myers Squibb, Roche and Novartis, is member of the scientific board of the Norwegian MS society, and has received financial support from the Research Council of Norway (grant #250864). RM has served on scientific advisory boards for Novartis Norway and Merck and received travel funding and/or speaker honoraria from Biogen, Novartis and Sanofi Genzyme. AE has received speaker honoraria from Biogen, Merck, Sanofi and Novartis. RE has received speaker honoraria from Novartis. HFH has received speaker honoraria from Biogen, Sanofi-Aventis, Merck, Novartis, and Roche. NØ has received speaker honoraria from Biogen, participated in clinical trials for Biogen and Sanofi-Aventis, and has served on a scientific advisory board for Novartis. FB has received compensation for steering/safety committee, activities and consulting services from Roche, Biogen, Merck, Combinostics, Janssen and IXICO. He is co-founder and shareholder of Queen Square Analytics. CET has a collaboration contract with ADx Neurosciences, Quanterix and Eli Lilly, performed contract research or received grants from AC-Immune, Axon Neurosciences, Biogen, Brainstorm Therapeutics, Celgene, EIP Pharma, Eisai, PeopleBio, Roche, Toyama, Vivoryon. She serves on editorial boards of Medidact Neurologie/Springer, Alzheimer Research and Therapy, Neurology: Neuroimmunology & Neuroinflammation, and is editor of a Neuromethods book Springer. LB has received unrestricted research grants to his institution and/or scientific advisory board or speaker honoraria from Almirall, Biogen, Genzyme, Merck, Novartis, Roche and Teva; and has participated in clinical trials organized by Biogen, Merck, Novartis, Roche, and Genzyme. ØT has received research grants and speaker honoraria from Biogen, Roche, Novartis, Merck and Sanofi. KMM has received unrestricted research grants to his institution; scientific advisory board, or speaker honoraria from Almirall, Biogen, Genzyme,

Merck, Novartis, Roche, and Teva; and has participated in clinical trials organized by Biogen, Merck, Novartis, and Roche. HV has received research grants from Pfizer, Merck Serono, Novartis and Teva; speaker honoraria from Novartis; and consulting fees from Merck Serono; all funds were paid directly to his institution.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Regional Committee for Medical and Health Research Ethics in Western Norway Regional Health Authority (OFAMS Study: clinical trials.gov, Identifier: NCT00360906). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iDs
Ingrid Anne Lie http://orcid.org/0000-0002-9948-2411
Sezgi Kaçar http://orcid.org/0000-0002-6695-2667
Rune Midgard http://orcid.org/0000-0002-6096-726X
Christian A Vedeler http://orcid.org/0000-0002-0993-8906

REFERENCES

1 Kutzelnigg A, Lucchinetti CF, Stadelmann C, *et al*. Cortical demyelination and diffuse white matter injury in multiple sclerosis. *Brain* 2005;128:2705–12.
2 Charil A, Zijdenbos AP, Taylor J, *et al*. Statistical mapping analysis of lesion location and neurological disability in multiple sclerosis: application to 452 patient data sets. *Neuroimage* 2003;19:532–44.
3 Steenwijk MD, Geurts JJG, Daams M, *et al*. Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. *Brain* 2016;139:115–26.
4 Ceccarelli A, Rocca MA, Pagani E, *et al*. A voxel-based morphometry study of grey matter loss in MS patients with different clinical phenotypes. *Neuroimage* 2008;42:315–22.
5 Comabella M, Montalban X. Body fluid biomarkers in multiple sclerosis. *Lancet Neurol* 2014;13:113–26.
6 Teunissen CE, Khalil M. Neurofilaments as biomarkers in multiple sclerosis. *Mult Scler* 2012;18:552–6.
7 Bielekova B, McDermott MP. Will CSF biomarkers guide future therapeutic decisions in multiple sclerosis? *Neurology* 2015;84:1620–1.
8 Kuhle J, Barro C, Disanto G, *et al*. Serum neurofilament light chain in early relapsing remitting MS is increased and correlates with CSF levels and with MRI measures of disease severity. *Mult Scler* 2016;22:1550–9.
9 Rosso M, Gonzalez CT, Healy BC, *et al*. Temporal association of sNfL and gad-enhancing lesions in multiple sclerosis. *Ann Clin Transl Neurol* 2020;7:945–55.
10 Bittner S, Oh J, Havrdová EK, *et al*. The potential of serum neurofilament as biomarker for multiple sclerosis. *Brain* 2021;144:2954–63.
11 Khalil M. Are neurofilaments valuable biomarkers for long-term disease prognostication in MS? *Mult Scler* 2018;24:1270–1.
12 Bridel C, Leurs CE, van Lierop ZYGJ, *et al*. Serum neurofilament light association with progression in natalizumab-treated patients with relapsing-remitting multiple sclerosis. *Neurology* 2021;97:e1898–905.
13 Cantó E, Barro C, Zhao C, *et al*. Association between serum neurofilament light chain levels and long-term disease course among patients with multiple sclerosis followed up for 12 years. *JAMA Neurol* 2019;76:1359–66.
14 Thebault S, Abdoli M, Fereshtehnejad S-M, *et al*. Serum neurofilament light chain predicts long term clinical outcomes in multiple sclerosis. *Sci Rep* 2020;10:10381.
15 Barro C, Benkert P, Disanto G, *et al*. Serum neurofilament as a predictor of disease worsening and brain and spinal cord atrophy in multiple sclerosis. *Brain* 2018;141:2382–91.
16 Chitnis T, Gonzalez C, Healy BC, *et al*. Neurofilament light chain serum levels correlate with 10-year MRI outcomes in multiple sclerosis. *Ann Clin Transl Neurol* 2018;5:1478–91.
17 Jakimovski D, Kuhle J, Ramanathan M, *et al*. Serum neurofilament light chain levels associations with gray matter pathology: a 5-year longitudinal study. *Ann Clin Transl Neurol* 2019;6:1757–70.
18 Filippi P, Vestenická V, Siarnik P, *et al*. Neurofilament light chain and MRI volume parameters as markers of neurodegeneration in multiple sclerosis. *Neuro Endocrinol Lett* 2020;41:17–26.
19 Tsagkas C, Parmar K, Pezold S, *et al*. Classification of multiple sclerosis based on patterns of CNS regional atrophy covariance. *Hum Brain Mapp* 2021;42:2399–415.
20 Malmeström C, Haghighi S, Rosengren L, *et al*. Neurofilament light protein and glial fibrillary acidic protein as biological markers in MS. *Neurology* 2003;61:1720–5.
21 Torkildsen O, Wergeland S, Bakke S, *et al*. ω-3 fatty acid treatment in multiple sclerosis (OFAMS Study): a randomized, double-blind, placebo-controlled trial. *Arch Neurol* 2012;69:1044–51.
22 Wesnes K, Myhr K-M, Riise T, *et al*. Low vitamin D, but not tobacco use or high BMI, is associated with long-term disability progression in multiple sclerosis. *Mult Scler Relat Disord* 2021;50:102801.
23 Varhaug KN, Barro C, Bjørnevik K, *et al*. Neurofilament light chain predicts disease activity in relapsing-remitting MS. *Neurol Neuroimmunol Neuroinflamm* 2018;5:e422.
24 Lie IA, Kerklingh E, Wesnes K, *et al*. The effect of gadolinium-based contrast-agents on automated brain atrophy measurements by FreeSurfer in patients with multiple sclerosis. *Eur Radiol* 2022;32:3576–87.
25 Schmidt P, Gaser C, Arsic M, *et al*. An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. *Neuroimage* 2012;59:3774–83.
26 Benkert P, Meier S, Schaedelin S, *et al*. Serum neurofilament light chain for individual prognostication of disease activity in people with multiple sclerosis: a retrospective modelling and validation study. *Lancet Neurol* 2022;21:246–57.
27 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 1995;57:289–300.
28 Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. *Nat Rev Immunol* 2015;15:545–58.
29 Vidal-Jordana A, Sastre-Garriga J, Pérez-Miralles F, *et al*. Early brain pseudoatrophy while on natalizumab therapy is due to white matter volume changes. *Mult Scler* 2013;19:1175–81.
30 Srpova B, Uher T, Hrnciarova T, *et al*. Serum neurofilament light chain reflects inflammation-driven neurodegeneration and predicts delayed brain volume loss in early stage of multiple sclerosis. *Mult Scler* 2021;27:52–60.
31 Sherman SM. Functioning of circuits connecting thalamus and cortex. *Compr Physiol* 2017;7:713–39.
32 Hasan KM, Kamali A, Kramer LA. Mapping the human brain white matter tracts relative to cortical and deep gray matter using diffusion tensor imaging at high spatial resolution. *Magn Reson Imaging* 2009;27:631–6.
33 Henry RG, Shieh M, Amirbekian B, *et al*. Connecting white matter injury and thalamic atrophy in clinically isolated syndromes. *J Neurol Sci* 2009;282:61–6.
34 Bergsland N, Laganà MM, Tavazzi E, *et al*. Corticospinal tract integrity is related to primary motor cortex thinning in relapsing-remitting multiple sclerosis. *Mult Scler* 2015;21:1771–80.
35 Meyer-Moock S, Feng Y-S, Maeurer M, *et al*. Systematic literature review and validity evaluation of the expanded disability status scale (EDSS) and the multiple sclerosis functional composite (MSFC) in patients with multiple sclerosis. *BMC Neurol* 2014;14:58.
36 Río J, Castilló J, Rovira A, *et al*. Measures in the first year of therapy predict the response to interferon beta in MS. *Mult Scler* 2009;15:848–53.
37 Sormani MP, Freedman MS, Aldridge J, *et al*. MAGNIMS score predicts long-term clinical disease activity-free status and confirmed disability progression in patients treated with subcutaneous interferon beta-1a. *Mult Scler Relat Disord* 2021;49:102790.

Appendix 1

Lesion segmentation. Lesion segmentation was done on FLAIR images using lesion segmentation tool (LST) (version 2.0.15; http://applied-statistics.de/lst.html)[1]. To optimise lesion filling, gadolinium-enhancing regions (both lesions and other regions) were first removed, by applying an upper intensity threshold at the 98th percentile. Next, the FMRIB Software Library (FSL) (version 5.0.10; http://www.fmrib.ox.ac.uk/fsl) was used to fill in abnormal voxels in these preprocessed T1-weighted images using the lesion_filling tool[2]. Then only the filled lesion voxels were pasted back into the original post-contrast 3D T1-weighted images to create the final lesion filled images.

Morphological reconstruction. Cortical reconstruction and parcellation for (local) cortical volume and thickness measurement and subcortical segmentation were performed with FreeSurfer version 7.1.1, a freely available software package for academic use, available through online download (http://surfer.nmr.mgh.harvard.edu/). The technical details of FreeSurfer procedures have been previously described[3, 4].

Quality control was performed by visual inspection, discarding cases with large segmentation errors. In cases where only specific anatomical regions were incorrectly segmented, we chose not to apply any manual corrections for these errors in our analyses.

The Desikan-Killiany atlas[5] was used to extract cortical thickness measures (mean cortical thickness, left and right hemisphere). Furthermore, total cerebral GM and WM volume, total deep GM and thalamus volume (left and right hemisphere) were obtained. Because of frequent suboptimal segmentation of the temporal pole, this region was excluded when calculating the total GM volume.

1. Schmidt P, Gaser C, Arsic M, et al. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *NeuroImage* 2012;59(4):3774-83.
2. Battaglini M, Jenkinson M, De Stefano N. Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Hum Brain Mapp* 2012;33(9):2062-71.
3. Fischl B. FreeSurfer. *NeuroImage* 2012;62(2):774-81.
4. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* 1999;9(2):179-94.
5. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 2006;31(3):968-80.

Supplemental material

Supplemental Table 1. The association of mean inflammatory sNfL level calculated from at least two measurements and fGd+ with MRI atrophy and clinical measures at the 10-year follow-up, corrected for age, sex, DMT use, eTIV (MRI volume measures), BL T2LC and BL Gd+LC.

Supplemental Table 1. Model 1: The association of inflammatory sNfL level and fGd+ with MRI atrophy and clinical measures at the 10-year follow-up, corrected for age, sex, DMT use, eTIV, BL T2LC and BL Gd+LC[a].

		Mean inflammatory sNfL level				**fGd+**				**Full model**	
MRI/clinical measure	*N*	*B*	*Std. B*	*95% conf. interval*	*p-value**	*B*	*Std. B*	*95% conf. interval*	*p-value**	R^2 *adj.*	*p-value*
Total GM volume	35				NS				NS		
Total WM volume	35				NS				NS		
Total subcortical GM volume	35	-96.27	-0.397	-175.064, -17.482	**0.036**				NS	0.132	0.018
Thalamus volume	35				NS				NS		
Mean Cth	35	-0.002	-0.433	-0.004, -0.001	**0.036**				NS	0.163	0.009
logLesion volume[b]	39				NS				NS		
logLesion count[b]	39	0.004	0.508	0.002, 0.007	**0.008**				NS	0.238	0.001
EDSS≥4[c]	45				NS				NS		
logT25FW[b]	43				NS				NS		
logChange in T25FW[b]	41				NS				NS		
logD9-HPT[b]	43				NS				NS		
logChange in D9-HPT[b]	41				NS	-0.554	-0.406	-0.928, -0.180	**0.010**	0.285	0.004

logND9-HPT[b]	42				NS				NS		
logChange in ND9-HPT[b]	41	-0.454	-0.320	-0.803, -0.105	**0.012**	-0.005	-0.405	-0.009, -0.001	**0.035**	0.367	0.001
PASAT	42				NS				NS		
Change in PASAT	40				NS				NS		
Oral SDMT	42				NS				NS		

[a] *Non-significant covariates removed from final model by backward elimination.*

[b] *Dependent variable log transformed due to non-normality (log-linear transformation).*

[c] *Analysed by logistic regression.*

**Adjusted p-values after controlling the False discovery rate (FDR) for multiple hypothesis testing.*

Abbreviations: sNfL=serum neurofilament light chain, eTIV=estimated total intracranial volume, DMT=disease modifying therapy, BL=baseline, Gd+=gadolinium-enhancing, LC=lesion count, fGd+=fraction of MRI scans with new Gadolinium-enhancing lesion, N=number, B=beta, Std=standardised, GM=grey matter, WM=white matter, NS=not significant, Cth=cortical thickness, EDSS=expanded disability status scale, T25FW=timed 25-foot walk, D9-HPT=dominant hand 9-hole peg test, ND9-HPT=non-dominant hand 9-hole peg test, PASAT=paced auditory serial addition test, SDMT=symbol digit modalities test.

Supplemental Table 2. The association of mean non-inflammatory sNfL level calculated from at least two measurements with MRI atrophy and clinical measures at the 10-year follow-up, corrected for age, sex, DMT use, eTIV (MRI volume measures), BL T2LC and BL Gd+LC.

Supplemental Table 2. Model 2: The association of mean non-inflammatory sNfL level with MRI atrophy and clinical measures at the 10-year follow-up, corrected for age, sex, DMT use, eTIV, BL T2LC and BL Gd+LC[a].

		Mean non-inflammatory sNfL level				**Full model**	
MRI/ clinical measure	*N*	*B*	*Std. B*	*95% conf. interval*	*p-value**	*R^2 adj.*	*p-value*
Total GM volume	59				NS		
Total WM volume	59				NS		
Total deep GM volume	59				NS		
Thalamus volume	59				NS		
Mean Cth	59				NS		
logLesion volume[b]	60				NS		
Lesion count	60				NS		
EDSS≥4[c]	67				NS		
logT25FW[b]	64				NS		
logChange in T25FW[b]	62				NS		
logD9-HPT[b]	63				NS		
logChange D9-HPT[b]	61				NS		
logND9-HPT[b]	62				NS		
Change in ND9-HPT	60				NS		
PASAT	65				NS		
Change in PASAT	63				NS		
Oral SDMT	60	0.446	0.389	0.168, 0.724	**0.002**	0.321	<0.001

[a] *Non-significant covariates removed from final model by backward elimination.*

[b] *Dependent variable log transformed due to non-normality (log-linear transformation).*

[c] *Analysed by logistic regression.*

**Adjusted p-values after controlling the False discovery rate (FDR) for multiple hypothesis testing.*

Abbreviations: sNfL=serum neurofilament light chain, eTIV=estimated total intracranial volume, DMT=disease modifying therapy, BL=baseline, Gd+=gadolinium-enhancing, LC=lesion count, N=number, B=beta, Std=standardised, GM=grey matter, WM=white matter, NS=not significant, Cth=cortical thickness, EDSS=expanded disability status scale, T25FW=timed 25-foot walk, D9-HPT=dominant hand 9-hole peg test, ND9-HPT=non-dominant hand 9-hole peg test, PASAT=paced auditory serial addition test, SDMT=symbol digit modalities test.

Supplemental table 3. The association of mean inflammatory sNfL level calculated from at least three measurements and fGd+ with MRI atrophy and clinical measures at the 10-year follow-up, corrected for age, sex, DMT use, eTIV (MRI volume measures), BL T2LC, BL Gd+LC and MRI atrophy measurements obtained at month 24.

Supplemental table 3. The association of inflammatory sNfL level and fGd+ with MRI atrophy measurements at the 10-year follow-up, corrected for age, sex, DMT use, eTIV, BL T2LC, Gd+LC, and MRI measurements at month 24[a]

		Mean inflammatory sNfL level				**fGd+**			**Full model**		
MRI measure	*N*	*B*	*Std. B*	*95% conf. interval*	*p-value**	*B*	*Std. B*	*95% conf. interval*	*p-value**	*R^2 adj.*	*p-value*
Total GM volume	19				NS				NS		
Total WM volume	19				NS				NS		
Total deep GM volume	19				NS				NS		
Thalamus volume	19				NS				NS		
Mean Cth	19				NS				NS		

[a] *Non-significant covariates removed from final model by backward elimination.*

**Adjusted p-values after controlling the False discovery rate (FDR) for multiple hypothesis testing.*

Abbreviations: sNfL=serum neurofilament light chain, eTIV=estimated total intracranial volume, DMT=disease modifying therapy, BL=baseline, Gd+=gadolinium-enhancing, LC=lesion count, fGd+=fraction of MRI scans with new Gadolinium-enhancing lesion, N=number, B=beta, Std=standardised, GM=grey matter, WM=white matter, NS=not significant, Cth=cortical thickness.

Supplemental Table 4. Multiple linear regression with cortical thickness of the left and right precentral gyrus as dependent variables, mean inflammatory sNfL level, D9-HPT, change in D9-HPT, ND9-HPT and change in ND9-HPT are included as independent variables in separate models, all corrected for age, sex, DMT use, BL T2LC, BL Gd+LC and fGd+.

Supplemental Table 4. The association of mean inflammatory sNfL level, D9-HPT and ND9-HPT with mean cortical thickness of the precentral gyri at the 10-year follow-up, corrected for age, sex, DMT use, BL T2LC, BL Gd+LC and fGd+[a].

Independent variable	*N*	*B*	*Std. B*	*95% conf. interval*	*p-value**	R^2 *adj.*	*p-value*
Cortical thickness left precentral gyrus							
Mean inflammatory sNfL level	25	-0.006	-0.498	-0.010, -0.001	**0.015**	0.216	0.011
D9-HPT	66	-0.011	-0.426	-0.018, -0.005	**0.001**	0.169	<0.001
Change in D9-HPT	62				NS		
ND9-HPT	64				NS		
Change in ND9-HPT	62				NS		
Cortical thickness right precentral gyrus							
Mean inflammatory sNfL level	25	-0.006	-0.488	-0.010, -0.001	**0.013**	0.206	0.013
D9-HPT	66	-0.010	-0.410	-0.016, -0.005	**0.001**	0.276	<0.001
Change in D9-HPT	62				NS		
ND9-HPT	64				NS		
Change in ND9-HPT	62				NS		

[a] *Non-significant covariates removed from final model by backward elimination.*

**Adjusted p-values after controlling the False discovery rate (FDR) for multiple hypothesis testing.*

Abbreviations: sNfL=serum neurofilament light, BL=baseline, Gd+=gadolinium-enhancing, LC=lesion count, fGd+=fraction of MRI scans with new Gadolinium-enhancing lesion, D9-HPT=dominant hand 9-hole peg test, ND9-HPT=non-dominant hand 9-hole peg test, DMT=disease modifying therapy, N=number, B=beta, Std=standardised, conf=confidence, adj=adjusted.

IV

RESEARCH ARTICLE OPEN ACCESS

The Effect of Smoking on Long-term Gray Matter Atrophy and Clinical Disability in Patients with Relapsing-Remitting Multiple Sclerosis

Ingrid Anne Lie, MD, Kristin Wesnes, MD, PhD, Silje S. Kvistad, MD, PhD, Iman Brouwer, MASc, Stig Wergeland, MD, PhD, Trygve Holmøy, MD, PhD, Rune Midgard, MD, PhD, Alla Bru, MD, Astrid Edland, MD, Randi Eikeland, MD, PhD, Sonia Gosal, MD, Hanne F. Harbo, MD, PhD, Grethe Kleveland, MD, Yvonne S. Sørenes, MD, Nina Øksendal, MD, Frederik Barkhof, MD, PhD, Hugo Vrenken, PhD, IR, Kjell-Morten Myhr, MD, PhD, Lars Bø, MD, Phd, and Øivind Torkildsen, MD, PhD

Correspondence
Dr. Lie
ingrid.lie@uib.no

Neurol Neuroimmunol Neuroinflamm 2022;9:e200008. doi:10.1212/NXI.0000000000200008

Abstract

Background and Objectives

The relationship between smoking, long-term brain atrophy, and clinical disability in patients with multiple sclerosis (MS) is unclear. Here, we assessed long-term effects of smoking by evaluating MRI and clinical outcome measures after 10 years in smoking and nonsmoking patients with relapsing-remitting MS (RRMS).

Methods

We included 85 treatment-naive patients with RRMS with recent inflammatory disease activity who participated in a 10-year follow-up visit after a multicenter clinical trial of 24 months. Smoking status was decided for each patient by 2 separate definitions: by serum cotinine levels measured regularly for the first 2 years of the follow-up (during the clinical trial) and by retrospective patient self-reporting. At the 10-year follow-up visit, clinical tests were repeated, and brain atrophy measures were obtained from MRI using FreeSurfer. Differences in clinical and MRI measurements at the 10-year follow-up between smokers and nonsmokers were investigated by 2-sample *t* tests or Mann-Whitney tests and linear mixed-effect regression models. All analyses were conducted separately for each definition of smoking status.

Results

After 10 years, smoking (defined by serum cotinine levels) was associated with lower total white matter volume ($\beta = -21.74$, $p = 0.039$) and higher logT2 lesion volume ($\beta = 0.22$, $p = 0.011$). When defining smoking status by patient self-reporting, the repeated analyses found an additional association with lower deep gray matter volume ($\beta = -2.35$, $p = 0.049$), and smoking was also associated with a higher score (higher walking impairment) on the log timed 25-foot walk test ($\beta = 0.050$, $p = 0.039$) after 10 years and a larger decrease in paced auditory serial addition test (attention) scores ($\beta = -3.58$, $p = 0.029$).

From the Department of Clinical Medicine (I.A.L., K.-M.M., L.B., Ø.T.), University of Bergen; Neuro-SysMed, Department of Neurology, Haukeland University Hospital (I.A.L., K.W., S.S.K., S.W., K.-M.M., Ø.T.), Bergen; St. Olav's University Hospital (K.W.), Trondheim; Department of Immunology and Transfusion Medicine (S.S.K.), Haukeland University Hospital, Bergen, Norway; Department of Radiology and Nuclear Medicine (I.B., F.B., H.V.), MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, location VUmc, The Netherlands; Norwegian Multiple Sclerosis Registry and Biobank (S.W.), Department of Neurology, Haukeland University Hospital, Bergen; Institute of Clinical Medicine (T.H., H.F.H.), University of Oslo; Department of Neurology, Akershus University Hospital (T.H.), Lørenskog; Department of Neurology (R.M.), Molde Hospital; Department of Neurology (A.B.), Stavanger University Hospital; Department of Neurology (A.E.), Vestre Viken Hospital Trust, Drammen; Department of Research and Education (R.E.), Sørlandet Hospital Trust, Kristiansand; Faculty of Health and Sport Science (R.E.), University of Agder, Grimstad; Department of Neurology (S.G.), Østfold Hospital Kalnes, Grålum; Department of Neurology (H.F.H.), Oslo University Hospital; Department of Neurology (G.K.), Innlandet Hospital Lillehammer; Department of Neurology (Y.S.S.), Haugesund Hospital; Department of Neurology (N.Ø.), Nordland Hospital Trust, Bodø, Norway; Institutes of Neurology and Healthcare Engineering (F.B.), University College London, Great Britain; and Norwegian Multiple Sclerosis Competence Centre (L.B.), Department of Neurology, Haukeland University Hospital, Bergen, Norway.

Go to Neurology.org/NN for full disclosures. Funding information is provided at the end of the article.

The Article Processing Charge was funded by the authors.

Glossary

BBB = blood-brain barrier; **BL** = baseline; **EDSS** = Expanded Disability Status Scale; **eTIV** = estimated total intracranial volume; **GM** = gray matter; **MS** = multiple sclerosis; **PASAT** = paced auditory serial addition test; **RRMS** = relapsing-remitting MS; **SPSS** = Statistical Product and Service Solutions; **T25FW** = timed 25-foot walk test; **VD** = vascular-disease; **WM** = white matter.

Discussion

Smoking was associated with brain atrophy and disability progression 10 years later in patients with RRMS. The findings imply that patients should be advised and offered aid in smoking cessation shortly after diagnosis, to prevent long-term disability progression.

Smoking is a known negative prognostic factor in patients with multiple sclerosis (MS), associated with higher disability[1] and higher risk of conversion to progressive disease phenotypes.[1,2] The causal relationship between these associations is not clear, and several pathophysiologic mechanisms have been proposed, for example: cigarette smoke triggering a proinflammatory cascade, inducing autoimmunity and heightened inflammatory activity; facilitating entry of immune cells to the CNS by disruption of the blood-brain barrier (BBB); epigenetic changes; and direct neurotoxicity due to mitochondrial damage.[3,4]

Studies examining the relation between smoking and inflammatory disease activity have reported inconsistent findings. Although some studies found smoking to be associated with higher relapse rates[5] and lesion loads,[1,6,7] 2 studies using cotinine, an alkaloid metabolite of nicotine,[8] to define smoking status, did not.[9,10]

In MS, smoking has cross-sectionally been shown to be associated with lower total brain volume,[1,6,11-13] but longitudinal relationships have been less well studied and with varying results.[1,7,9] Furthermore, few studies have considered the relation between smoking and gray matter (GM) atrophy, again with conflicting results.[6,11,13]

To better understand the possible adverse prognostic effects of smoking in MS, we aimed to investigate the relation between smoking and long-term brain lesion load, atrophy, and clinical outcome measures, first by comparing smokers and nonsmokers, defined by both patient self-reporting and cotinine levels, and second by studying a possible dose effect, using mean cotinine levels in smoking patients with MS.

Methods

Participants

The patients included in this study participated in a 10-year follow-up visit, after a multicenter trial on ω-3 fatty acids in MS (the OFAMS-study). In the original trial, a total of 92 patients with relapsing-remitting MS (RRMS) were followed up for 24 months; a detailed description of the study is provided elsewhere.[14] During the OFAMS study period, patients attended regular follow-up visits at their local study site, undergoing biochemical, radiologic, and clinical examinations, including the Expanded Disability Status Scale (EDSS), timed 25-foot walk test (T25FW), the dominant hand and nondominant hand 9-hole peg test, and the paced auditory serial addition test (PASAT). Ten years after the trial concluded, all available (87) participants were invited to a follow-up visit, of which 85 accepted.[15] At the 10-year follow-up visit, the patients repeated the radiologic and clinical examinations. Between the OFAMS-study and the 10-year follow-up visit, the participants had received routine clinical treatment and care.

Standard Protocol Approvals, Registrations, and Patient Consents

The OFAMS-study and the 10-year follow-up were approved by the Regional Committee for Medical and Health Research Ethics in Western Norway Regional Health Authority (OFAMS-study: clinicaltrials.gov, Identifier: NCT00360906). All participants gave their written informed consent.

Cotinine Measurement

Serum samples were stored at −80°C until analysis and performed simultaneously for all samples from each patient. As previously described,[10] serum cotinine levels were measured by liquid chromatography tandem mass spectrometry at Bevital AS (Bergen, Norway). Laboratory technicians were blinded to patient clinical status. Serum cotinine levels were analyzed from samples collected during the OFAMS-study at baseline (BL) and months 6, 12, 18, and 24.

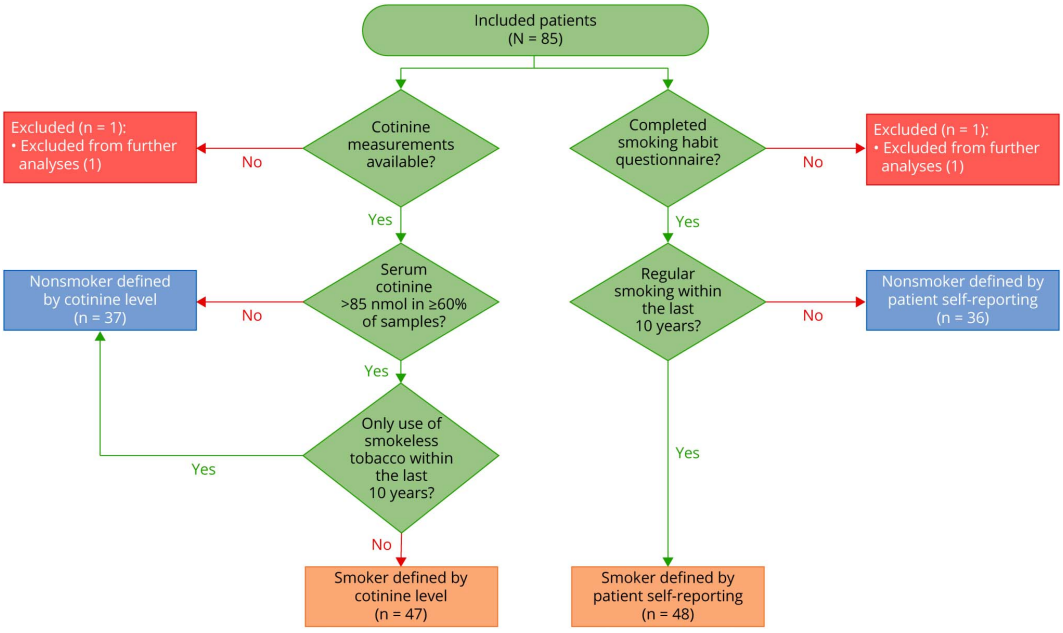
Definitions of Smokers and Nonsmokers

The effect of smoking status on long-term MRI and clinical outcome measures was analyzed using 2 definitions (described below) of smoking status separately: by (1) serum cotinine levels, with the findings presented in the main text and (2) patient self-reporting, with the findings presented in the supplemental material.

Smoking Defined by Cotinine Levels

Cotinine levels >85 nmol/L indicate recent tobacco use[16] and are regarded to distinguish tobacco users from nontobacco users in the general population.[17] Patients were categorized into 2 groups according to serum cotinine level: smokers were defined as patients with serum cotinine level >85 nmol/L in ≥60% of the samples and nonsmokers were defined as patients with serum cotinine levels ≤85 nmol/L in ≥60% of the samples. Based on previous studies not finding nicotine or smokeless tobacco to be associated with MS disease progression,[18,19] patients who at the 10-year follow-up visit

Figure 1 Flowchart Illustrating the Classification Process According to the 2 Definitions of Smoking



reported "snuff" use (oral tobacco), but no smoking for the past 10 years, were defined as nonsmokers (Figure 1).

Smoking Defined by Patient Self-Reporting

At the 10-year follow-up visit, patients answered a questionnaire about habits of tobacco use, including smoking and snuff use. Patients who reported to have smoked regularly within the past 10 years were defined as smokers while patients who did not report regular smoking, or reported only snuff use (6 patients), were defined as nonsmokers (Figure 1).

Risk Factors and Presence of Peripheral, Cardiovascular, or Cerebrovascular Disease

Vascular risk factors and established peripheral, cardiovascular, and cerebrovascular disease are independently associated with brain imaging changes.[20] As smoking is a known risk factor for such conditions,[21] patient-reported hypertension, dyslipidemia, hypercoagulable disorder, and symptomatic cardiovascular or cerebrovascular disease and/or events was designated as a dichotomous vascular disease (VD) factor and included as a covariate in the final analyses.

MRI Data and Analysis

The 10-year Follow-up Visit

Imaging was performed at the different study sites, on a 3 Tesla (T) MRI scanner if available, alternatively using a 1.5T MRI scanner, with a standard head coil. The acquisition protocol included the following MRI sequences: a T2-weighted 3D sagittal fluid attenuated inversion recovery (FLAIR) (resolution: 1 × 1 × 1 mm3, echo time (TE)/repetition time (TR)/inversion time (TI) = 386/5000/1.65–2.2 ms) and a postcontrast T1-weighted 3D sagittal magnetization-prepared rapid gradient echo sequence (resolution: 1 × 1 × 1mm3, TE/TR/TI = 2.28/1800/900 ms, flip angle 8°). Acquisition details across sites are presented in eTable 1, links.lww.com/NXI/A728.

Lesion Segmentation

Lesion segmentation was performed on FLAIR images using lesion segmentation tool (version 2.0.15; applied-statistics.de/lst.html).[22] To optimize lesion filling, gadolinium-enhancing regions (both lesions and other regions) were first removed, by applying an upper intensity threshold at the 98th percentile. Next, the FMRIB Software Library (version 5.0.10[23]) was used to fill in abnormal voxels in these preprocessed T1-weighted images using the lesion_filling tool.[24] Then, only the filled lesion voxels were pasted back into the original postcontrast 3D T1-weighted images to create the final lesion filled images.

Morphological Reconstruction

Cortical reconstruction and parcellation for (local) cortical volume and thickness measurement and subcortical segmentation were performed with FreeSurfer version 7.1.1, a freely available software package for academic use, available through online download.[25] The technical details of FreeSurfer procedures have been previously described.[26,27] The use of FreeSurfer on postcontrast 3D T1-weighted images as applied here, was recently validated.[28]

Table 1 Demographic and Clinical Characteristics

	BL	10-y follow-up visit	BL	10-y follow-up visit
Smoking status defined by cotinine level	Nonsmokers (37)		Smokers (47)	
Age in y, mean (SD)	37.5 (6.8)	49.4 (6.9)	37.9 (9.7)	49.7 (9.7)
Sex, female, N (%)	22 (59.5)		32 (68.1)	
Time from diagnosis, mean in y (SD)	2.8 (3.5)	14.8 (3.6)	2.2 (3.0)	14.3 (3.1)
Disease phenotype (N)	RRMS (37)	RRMS (35), SPMS (2)	RRMS (47)	RRMS (42), SPMS (5)
EDSS, mean (SD)	1.8 (0.8)	2.6 (1.6)	2.0 (0.9)	2.9 (1.6)
Mean cotinine level (SD)[a]			850.6 (454.5)	
Smoking status defined by patient self-reporting	Nonsmokers (36)		Smokers (48)	
Age in y, mean (SD)	37.5 (7.6)	49.4 (7.7)	37.9 (9.2)	49.7 (9.2)
Sex, female, N (%)	22 (61.1)		33 (68.8)	
Time from diagnosis, mean in y (SD)	2.8 (3.6)	14.9 (3.7)	2.3 (3.0)	14.3 (3.1)
Disease phenotype (N)	RRMS (36)	RRMS (35), SPMS (1)	RRMS (48)	RRMS (42), SPMS (6)
EDSS, mean (SD)	1.8 (0.8)	2.6 (1.6)	2.0 (0.9)	2.9 (1.6)

Abbreviations: BL = baseline; EDSS = Expanded Disability Status Scale; RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary progressive multiple sclerosis.
[a] Mean serum cotinine level calculated from measurements obtained from BL to month 24.

Quality control was performed by visual inspection, discarding cases with large segmentation errors. Minor to moderate segmentation errors of specific anatomic regions were found in all scans and are previously shown to occur more frequently and to a more severe degree in postcontrast images.[28] As these errors were so commonly occurring, mostly with the same effect (overestimation of GM volume/cortical thickness[28]), we chose to not apply any manual corrections for these errors in our analyses.

The Desikan-Killiany atlas[29] was used to extract cortical thickness measures. The mean cortical thickness of the left and right hemisphere was averaged to calculate the overall mean cortical thickness. Furthermore, total cerebral GM and white matter (WM) volume and total deep GM and thalamus volume (average of left and right hemisphere) were obtained. Because of frequent suboptimal segmentation of the temporal pole (previously found in a minor to moderate degree in almost 50% of postcontrast scans[28]), this region was excluded when calculating the total GM volume.

Statistical Analysis

Statistical analyses were performed using Statistical Product and Service Solutions (SPSS) for macOS (Version 25; SPSS, Chicago, IL) and R software (V.4.0.5).

Kolmogorov-Smirnov tests and visual inspection of the histograms were used to assess the normality of the variables.

The primary outcome measures were MRI and clinical measurements obtained cross-sectionally at the 10-year follow-up visit and the difference in clinical measurements from month 24 to the 10-year follow-up.

As a first exploratory analysis, the difference in outcome measures between smokers and nonsmokers was analyzed by 2-sample *t* tests for normally distributed variables; otherwise, Mann-Whitney tests were used.

The relationship of smoking status and mean cotinine level with MRI and clinical outcome measures was then further investigated by a linear mixed-effect regression model, correcting for age, sex, vascular disease, estimated total intracranial volume (eTIV) (eTIV included as a covariate in analyses regarding brain volume measurements), BL EDSS, and time from diagnosis. To correct for scanner variability, MRI scanner was entered as a random effect. Assumptions for linear regression were checked for each final model; if the assumptions were not satisfied, log-linear transformation was performed (e.g., logT2 lesion volume). For variables with values below 1, a constant was added before log transformation.

Data Availability

Data not provided within this article may be shared (anonymized) by request from a qualified investigator.

Results

Patient Characteristics

We included the 85 patients who participated in the 10-year follow-up. Each patient was classified as a smoker or nonsmoker, by the 2 different definitions of smoking: (1) by cotinine levels measured during the OFAMS-study and (2) by retrospective patient self-reporting at the 10-year follow-up

Table 2 MRI and Clinical Measures in Smokers and Nonsmokers (Defined by Cotinine Level) at the 10-Year Follow-up Visit and Change in Clinical Measures Between Month 24 and the 10-Year Follow-up Visit

MRI/clinical measure	Nonsmokers (mean, SD)	Smokers (mean, SD)	Mean difference	95% CI	*p* Value
Total GM volume (mL)	643.62 (55.76)	621.83 (47.51)	21.785	−3.220, 46.791	0.087
Total WM volume (mL)	465.33 (54.98)	435.76 (42.33)	29.577	5.961, 53.192	**0.015**
Total deep GM volume (mL)	57.73 (4.98)	54.59 (5.50)	3.139	0.586, 5.691	**0.017**
Thalamus volume (mL)	8.08 (1.02)	7.67 (1.00)	0.41	−0.08, 0.90	0.096
Mean Cth (mm)	2.55 (0.13)	2.52 (0.12)	0.03	−0.03, 0.09	0.329
T2 lesion volume (mL)[a]	3.11 (3.07)	6.63 (9.68)			**0.014**
T2 lesion count	19.70 (7.61)	21.62 (9.85)	−1.918	−6.115, 2.278	0.365
EDSS[a]	2.5 (1.5)	2.0 (1.6)			0.306
Change in EDSS[a]	0.5 (2.0)	0.0 (1.13)			0.505
T25FW[a]	3.88 (1.29)	4.43 (1.47)			**0.031**
Change in T25FW[a]	−0.01 (1.30)	0.31 (1.23)			0.202
D9-HPT[a]	20.02 (6.59)	21.31 (5.68)			0.167
Change in D9-HPT[a]	2.41 (3.11)	3.11 (4.65)			0.480
ND9-HPT[a]	20.68 (6.59)	23.23 (7.27)			0.130
Change in ND9-HPT[a]	3.14 (4.80)	4.08 (6.11)			0.307
PASAT	49.06 (8.40)	45.11 (10.68)	3.95	−0.43, 8.32	0.077
Change in PASAT	−3.91 (7.20)	−7.26 (6.70)	3.35	0.13, 6.57	**0.042**

Abbreviations: Cth = cortical thickness; D9-HPT = dominant hand 9-hole peg test; EDSS = Expanded Disability Status Scale; GM = gray matter; ND9-HPT = nondominant hand 9-hole peg test; PASAT = paced auditory serial addition test; T25FW = timed 25-foot walk; WM = white matter.
[a] Difference analyzed by Mann-Whitney *U* test, median, and interquartile range reported.
Bold text indicates statistically significant *p* values.

(Figure 1). Samples available for cotinine analyses were missing for one patient, and another patient did not complete the questionnaire concerning tobacco use, leaving 84 patients to be classified as smoker or nonsmoker by each definition. Including the 2 patients missing either cotinine measurements or the questionnaire, 9 patients were classified differently based on the 2 definitions. Forty-seven patients were smokers and 37 nonsmokers defined by cotinine levels. By patient self-reporting, 48 patients were smokers and 36 nonsmokers. Of the 48 smokers defined by patient self-reporting, 47 reported to smoke 10 years ago. Additional patient self-reported smoking habits are listed in supplemental eTable 2, links.lww.com/NXI/A728.

The mean follow-up time from the BL and month 24 visit to the 10-year follow-up visit was 12.0 (±0.6) and 10.0 (±0.6) years, respectively. Table 1 summarizes the clinical characteristics of the included patients.

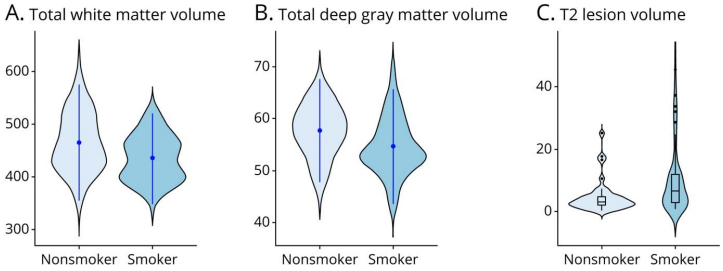
Difference in MRI and Clinical Outcome Measures Between Smokers and Nonsmokers

The results of the exploratory *t* tests are shown in Table 2 (smoking defined by cotinine levels) and supplemental eTable 3, links.lww.com/NXI/A728 (smoking defined by patient self-reporting). In brief, total WM ($p = 0.015$) and deep GM ($p = 0.017$) volumes were significantly smaller, and total T2 lesion volumes were significantly larger ($p = 0.014$) in smokers defined by cotinine level (figure 2). For the clinical measures, smokers had a higher score (more disability) on the T25FW test ($p = 0.031$) after 10 years and a larger decrease in attention scores measured by the PASAT test ($p = 0.042$) between month 24 and the 10-year follow-up visit (figure 3). The results were similar for smoking defined by patient self-reporting (eTable 3).

Smoking Status and Long-term MRI and Clinical Outcome Measures

The results from the linear mixed-model investigating the relationship between smoking and long-term MRI and clinical measures are shown in Table 3 (smoking defined by cotinine levels) and supplemental eTable 4, links.lww.com/NXI/A728 (smoking defined by patient self-reporting). Smoking defined by cotinine level was associated with lower total WM volume ($\beta = -21.74$, $p = 0.039$) after 10 years and with higher total logT2 lesion volume ($\beta = 0.22$, $p = 0.011$). Similar results were found for smoking defined by patient self-reporting (eTable 4), additionally associated with lower deep GM volume ($\beta = -2.35$, $p = 0.049$) after 10 years. There were no significant associations between smoking and clinical disability when

Figure 2 Distribution Plots of MRI Measurements at the 10-Year Follow-up Visit in Nonsmokers and Smokers



The width of the shaded area represents the proportion of observations for (A) total white matter volume (mL) and (B) total deep gray matter volume (mL) (point range represents the mean and SD), and (C) T2 lesion volume (mL) (box plots represent the median and IQR, with the whiskers representing the distribution of observations within x1.5 of the IQR). IQR = interquartile range.

defined by cotinine levels (Table 3), but when defined by patient self-reporting, smoking was associated with a higher score on the logT25FW test (β = 0.050, p = 0.039) after 10 years and a larger decrease in PASAT scores (β = −3.58, p = 0.029) (eTable 4).

Dose-Effect Relationship on Long-term MRI and Clinical Outcome Measures

The results from the mixed-effect model assessing the relation between mean cotinine levels in smokers (defined by cotinine level) and MRI and clinical outcome measures are shown in supplemental eTable 5, links.lww.com/NXI/A728. In a bivariate model, higher cotinine levels were significantly associated with lower mean cortical thickness, but not after adjusting for age, sex, VD, BL EDSS, and time from diagnosis.

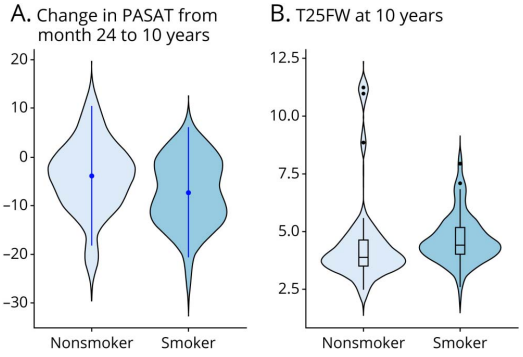
Discussion

We found that patients with MS who smoked had lower total WM and deep GM volumes and higher T2 lesion volumes after 10 years. Smokers also accumulated more disability and incurred a larger decrease in attention scores measured by the T25FW and PASAT test, respectively. The results obtained by defining smokers by cotinine levels collected during the initial 2-year period were highly comparable with those obtained when defining smokers by retrospective patient self-reporting. Furthermore, the associations were still significant when correcting for vascular risk factors and established cardiovascular disease. These findings suggest that smoking has a negative long-term influence on prognosis and disease progression in patients with MS.

The association between smoking and higher lesion load is partly in line with previous research,[1,7] supporting the notion that smoking could heighten inflammatory activity.[3] However, smoking and increased lesional activity have not been consistently related,[9,10] as also shown in a previous study investigating the same patient cohort as in this current work. In the previous study, no association was found between tobacco use (defined by cotinine levels) and the occurrence of new or enlarging lesions during the trial period of 24 months.[10] The discrepant findings in this same study population may be a result of the different outcome measures used, that is, lesion volumes vs the less sensitive dichotomous measure of new or expanding lesions present or not present, especially considering the more limited follow-up time of 2 years. Moreover, it is unknown how smoking may affect pathologic processes within the occurring lesions. Lesions with persistent subtle inflammation, called chronic active or smoldering lesions, are associated with low-grade BBB leakage,[31] higher atrophy rates,[32] and have been shown to develop in patients with RRMS and slowly expand over years.[33] In this study, we were unable to investigate whether smokers had a higher fraction of smoldering lesions, but testing that hypothesis could provide insight to the discrepant associations found for the occurrence of new lesions and total lesion volume and to the overall worse prognosis seen in smoking patients with MS.[3]

After 10 years, we found that deep GM and total WM volume were lower in patients who smoked. Previous longitudinal studies on smoking/nonsmoking patients with MS have mainly assessed whole-brain atrophy and have not reported consistent results.[1,7,9] Cross-sectional studies assessing GM atrophy showed a similar lack of consensus.[6,11,13] In studies investigating non-MS populations, smoking is associated with atrophy most evidently in the frontal and temporal lobe, cingulate gyrus, and the cerebellum[33,34] while the associations with subcortical GM are more variable.[33,34] Furthermore, neither studies on MS nor those on non-MS populations have found smoking to be related to lower WM volume.[6,7,11,13,33,34] In early MS, atrophy in the deep GM has previously been shown to develop at a relatively high rate[35,36] compared with other GM regions, and to be closely related to WM lesions.[36,37] Although the causal mechanisms are still not sufficiently clear, a spatiotemporal relationship between WM lesions and subsequent deep GM atrophy progression through neuroaxonal degradation[38] seems likely, explaining at least part of the neurodegenerative process. Together with the higher lesion volume, our findings of lower deep GM and total WM volumes in smoking patients with MS may suggest that the pathologic changes are driven by increased inflammatory damage in the WM, followed by secondary degeneration in regions either consisting of WM or highly connected through

Figure 3 Distribution Plots of Clinical Measurements in Nonsmokers and Smokers



The width of the shaded area represents the proportion of observations for (A) change in PASAT score from month 24 to 10 years (point range represents the mean and SD) and (B) T25FW test score at 10 years (box plots represent the median and IQR, with the whiskers representing the distribution of observations within x1.5 of the IQR). IQR = interquartile range; PASAT = paced auditory serial addition test; T25FW = timed 25-foot walk.

WM tracts.[39] The hypothesis that the main neurodegenerative pathway in MS is driven by neuroaxonal injury has also been proposed in a recent study, which observed higher levels of neurofilament light chain in smokers.[40]

Although not significant after correcting for age, sex, vascular risk factors, BL EDSS, and time since diagnosis, our initial bivariate analysis found that higher cotinine levels in smokers were associated with lower cortical thickness after 10 years. This is similar to previous literature on non-MS populations, finding a dose-dependent relationship between smoking and cortical atrophy.[34,41] The different smoking-associated atrophy patterns seen when comparing smoking and nonsmoking patients with MS (lower deep GM and total WM volume), and cotinine levels in smoking patients (lower cortical thickness), may suggest that smoking affects atrophy progression both through MS-specific disease mechanisms and directly in a dose-dependent manner. It is, however, important to highlight that our sample size for the dose-effect analyses was small and that the current data did not allow us to investigate these possible mechanisms directly.

At the 10-year follow-up visit, smokers had a higher degree of walking impairment measured by the T25FW test, indicating higher risk of disability accrual. The lack of association between smoking defined by cotinine levels and change in EDSS score over the follow-up has previously been reported in this current patient population.[15] In our study, we confirmed this finding, for both definitions of smoking. The results are partly in line with previous research, much of which reported a relation between smoking and higher EDSS or Multiple Sclerosis Severity Score in cross-sectional analyses,[1,6] but more variably longitudinally.[1,9,40,42,43] A possible cause of the inconsistent results for EDSS is the known low sensitivity to change for this measurement,[44] suggesting that more targeting tests should be used to capture disability progression in specific functions. Overall, the associations with disability accrual were modest in our study, and after correcting for relevant covariates, it was only found for one of the 2 definitions of smoking (patient self-reporting). In future studies, these longitudinal relations should be further investigated in larger patient populations.

In our study, patients who smoked also had a higher decrease in the PASAT score from month 24 to the 10-year follow-up. This finding is in line with previous studies, where smokers tended to perform worse cognitively.[40,45] Attention, information processing, and working memory are cognitive domains commonly affected in patients with MS,[46] but impairment on these domains has also been shown in smoking non-MS populations.[34] This may imply that smoking patients with MS are at additive risk of developing cognitive impairment, through mechanisms specific to both MS and smoking. This suggests an important clinical consequence to minimize long-term GM atrophy and clinical decline; people with MS who smoke should be encouraged and assisted to quit smoking.

Our study is not without limitations. Brain atrophy and lesion load were measured cross-sectionally at the 10-year follow-up visit. In future research, long-term measurements should be corrected for values at baseline if available, or be investigated by longitudinal analyses, to conclude on atrophy progression in smoking patients with MS.

Obtaining volume and cortical thickness measures in postcontrast images by FreeSurfer is not the standard approach. Recent work using data from a subgroup of this patient cohort has demonstrated excellent consistency between values obtained from precontrast and postcontrast 3D T1-weighted images,[28] although minor to moderate segmentation errors, especially in the temporal lobe were more common in postcontrast images. In this current study, we did not apply manual corrections for these errors, and the volume of the temporal lobe was excluded from the total GM volume to limit the possible bias introduced by the larger variability in measurements extracted from this region. In future studies, corrections of these segmentation errors should be considered, especially if evaluating regional atrophy measures.

Owing to its stability in plasma over time (half-life of approximately 20 hours), cotinine has become the preferred biomarker to quantify long-term nicotine exposure.[8] However, the use of cotinine levels as a proxy for smoking is potentially biased by other sources of nicotine. This is an important limitation because neither nicotine nor smokeless tobacco have shown to induce inflammation in MS[19] or increase the risk of the disease.[18] In our study, patients who at the 10-year follow-up reported use of smokeless tobacco (e.g., snuff) exclusively (4 patients) were, therefore, classified as nonsmokers, regardless of their measured cotinine levels. Furthermore, the results obtained by defining smoking by

Table 3 Effect of Smoking Status (Defined by Cotinine Level) on MRI and Clinical Measures at the 10-Year Follow-up Visit and Change in Clinical Measures Between Month 24 and the 10-Year Follow-up Visit

MRI/clinical measure	N	Beta	St. Error	95% CI	*p* Value
Total GM volume (mL)	67	−12.025	9.419	−29.349, 5.299	0.208
Total WM volume (mL)	67	−21.739	10.254	−40.599, −2.880	**0.039**
Total deep GM volume (mL)	67	−2.109	1.127	−4.183, −0.036	0.068
Thalamus volume (mL)	67	−0.352	0.221	−0.758, 0.054	0.118
Mean Cth (mm)	67	−0.031	0.025	−0.076, 0.015	0.214
LogT2 lesion volume (mL)[a]	64	0.224	0.084	0.069, 0.379	**0.011**
T2 lesion count	64	2.095	2.302	−2.163, 6.352	0.368
EDSS	67	0.140	0.289	−0.396, 0.676	0.630
Change in EDSS	65	−0.234	0.301	−0.792, 0.324	0.441
LogT25FW[a]	66	0.012	0.024	−0.032, 0.056	0.622
LogChange in T25FW[a]	64	0.013	0.020	−0.024, 0.051	0.511
LogD9-HPT[a]	64	0.035	0.023	−0.007, 0.078	0.130
LogChange in D9-HPT[a]	62	0.044	0.049	−0.047, 0.135	0.377
LogND9-HPT[a]	64	0.022	0.021	−0.017, 0.061	0.300
LogChange in ND9-HPT[a]	62	0.064	0.062	−0.051, 0.179	0.310
PASAT	64	−2.572	2.167	−6.581, 1.436	0.242
Change in PASAT	62	−2.654	1.565	−5.543, 0.236	0.097

Abbreviations: Cth = cortical thickness; D9-HPT = dominant hand 9 hole-peg test; EDSS = Expanded Disability Status Scale; GM = gray matter; ND9-HPT = nondominant hand 9-hole peg test; PASAT = paced auditory serial addition test; T25FW = timed 25-foot walk; WM = white matter.
[a] Dependent variable log transformed because of nonnormality (log-linear transformation).
Bold text indicates statistically significant *p* values.

cotinine levels were comparable with those using patient self-reporting to define smoking habits. The associations with MRI atrophy and clinical measures were somewhat stronger when using the definition based on patient self-reporting, most notably for clinical disability. This may be explained by this definition also capturing patients who smoked regularly in time periods during the follow-up, after the 2 years of the clinical trial. In this study, serum cotinine measurements were not available after the first 2 years of follow-up or at the 10-year follow-up visit. In future works, the effect of cotinine levels measured regularly over a longer follow-up period should be explored, especially to determine the effect of smoking duration and cessation. Nevertheless, the overall comparable results suggest that serum cotinine levels provide an objective and reliable option for defining smoking habits and especially to investigate dose-dependent relationships.

Several comorbid conditions may independently influence brain tissue changes (including gray and WM atrophy and localized WM hyperintensities)[20] and can be caused or exacerbated by smoking.[21] We attempted to limit the effect of these complex interrelations by correcting for vascular risk factors and established cardiovascular and cerebrovascular disease. MS disease–related factors, such as disease-modifying therapies, are also likely to affect brain atrophy. In this data set, however, this was not possible to statistically consider because the patients had used a variety of therapies at different times and duration over the follow-up.

Smoking was associated with lower deep GM and total WM volume and higher T2 lesion volume after 10 years in patients with RRMS. Patients who smoked had higher physical and cognitive disability accrual, measured by the T25FW and PASAT test, respectively. The findings suggest that smoking patients with MS should be advised and offered aid in smoking cessation as early as possible in the disease course.

Acknowledgment

The authors thank the OFAMS study group and the patients who participated in the study.

Study Funding

The authors report no targeted funding.

Disclosure

The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article. I.A. Lie has received research grants

from the Meltzer Research Fund, Gerda Meyer Nyquist Guldbrandson & Gert Meyer Nyquists Legat, and the Independent Order of Odd Fellows; K. Wesnes has received unrestricted research grants from Novartis and Biogen, research grant from the Independent Order of Odd Fellows, and speaker honoraria from Biogen; S.S. Kvistad has received unrestricted research grants from Novartis and Biogen; I. Brouwer has received research support from Merck KGaA, Novartis, and Teva; S. Wergeland has received speaker honoraria from and served on scientific advisory boards for Biogen, Janssen-Cilag, Sanofi, and Novartis; T. Holmøy has received speaker honoraria, research support/grants and participated in clinical trials for Biogen, Merck, Sanofi, Bristol Myers Squibb, Roche and Novartis, is member of the scientific board of the Norwegian MS society, and has received financial support from the Research Council of Norway (grant #250864); R. Midgard has served on scientific advisory boards for Novartis Norway and Merck and received travel funding and/or speaker honoraria from Biogen, Novartis, and Sanofi Genzyme; A. Bru reports no disclosures relevant to the manuscript; A. Edland has received speaker honoraria from Biogen, Merck, Sanofi, and Novartis; R. Eikeland has received speaker honoraria from Novartis; S. Gosal reports no disclosures relevant to the manuscript; H.F. Harbo has received speaker honoraria from Biogen, Sanofi-Aventis, Merck, Novartis, and Roche; G. Kleveland reports no disclosures relevant to the manuscript; Y.S. Sørenes reports no disclosures relevant to the manuscript; N. Øksendal has received speaker honoraria from Biogen, participated in clinical trials for Biogen and Sanofi-Aventis and has served on a scientific advisory board for Novartis; F. Barkhof has received compensation for steering/safety committee, activities and consulting services from Roche, Biogen, Merck, Combinostics, Janssen, and IXICO. He is cofounder and shareholder of Queen Square Analytics LTD; H. Vrenken has received research grants from Pfizer, Merck Serono, Novartis and Teva, speaker honoraria from Novartis, and consulting fees from Merck Serono. All funds were paid directly to his institution; K.M. Myhr has received unrestricted research grants to his institution, scientific advisory board, and speaker honoraria from Biogen, Sanofi, Merck, Novartis, and Roche and has participated in clinical trials organized by Biogen, Merck, Novartis, and Roche; L. Bø has received unrestricted research grants to his institution and/or scientific advisory board or speaker honoraria from Almirall, Biogen, Genzyme, Merck, Novartis, Roche, and Teva and has participated in clinical trials organized by Biogen, Merck, Novartis, Roche, and Genzyme; Ø. Torkildsen has received research grants and speaker honoraria from Biogen, Roche, Novartis, Merck, and Sanofi and has participated in clinical trials organized by Merck and Sanofi. Go to Neurology.org/NN for full disclosures.

Publication History

Received by *Neurology: Neuroimmunology & Neuroinflammation* January 26, 2022. Accepted in final form May 6, 2022. Submitted and externally peer reviewed. The handling editor was Friedemann Paul.

References

1. Healy BC, Ali EN, Guttmann CR, et al. Smoking and disease progression in multiple sclerosis. *Arch Neurol.* 2009;66(7):858-864.
2. Degelman ML, Herman KM. Smoking and multiple sclerosis: a systematic review and meta-analysis using the Bradford Hill criteria for causation. *Mult Scler Relat Disord.* 2017;17:207-216.
3. Rosso M, Chitnis T. Association between cigarette smoking and multiple sclerosis: a review. *JAMA Neurol.* 2020;77(2):245-253.
4. Wingerchuk DM. Smoking: effects on multiple sclerosis susceptibility and disease progression. *Ther Adv Neurol Disord.* 2011;5(1):13-22.
5. Petersen ER, Søndergaard HB, Laursen JH, et al. Smoking is associated with increased disease activity during natalizumab treatment in multiple sclerosis. *Mult Scler.* 2018;25(9):1298-1305.
6. Zivadinov R, Weinstock-Guttman B, Hashmi K, et al. Smoking is associated with increased lesion volumes and brain atrophy in multiple sclerosis. *Neurology.* 2009;73(7):504-510.
7. Horakova D, Zivadinov R, Weinstock-Guttman B, et al. Environmental factors associated with disease progression after the first demyelinating event: results from the multi-center SET study (P05.134). *Neurology.* 2013;80(7 suppl):P05134.
8. Florescu A, Ferrence R, Einarson T, Selby P, Soldin O, Koren G. Methods for quantification of exposure to cigarette smoking and environmental tobacco smoke: focus on developmental toxicology. *Ther Drug Monit.* 2009;31(1):14-30.
9. Munger KL, Fitzgerald KC, Freedman MS, et al. No association of multiple sclerosis activity and progression with EBV or tobacco use in BENEFIT. *Neurology.* 2015;85(19):1694-1701.
10. Kvistad S, Myhr KM, Holmøy T, et al. No association of tobacco use and disease activity in multiple sclerosis. *Neurol Neuroimmunol Neuroinflamm.* 2016;3(4):e260.
11. Kappus N, Weinstock-Guttman B, Hagemeier J, et al. Cardiovascular risk factors are associated with increased lesion burden and brain atrophy in multiple sclerosis. *J Neurol Neurosurg Psychiatry.* 2016;87(2):181-187.
12. Durhan G, Diker S, Has AC, 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):805-811.
13. Graetz C, Groger A, Luessi F, et al. Association of smoking but not HLA-DRB1*15:01, APOE or body mass index with brain atrophy in early multiple sclerosis. *Mult Scler.* 2019;25(5):661-668.
14. Torkildsen O, Wergeland S, Bakke S, et al. omega-3 fatty acid treatment in multiple sclerosis (OFAMS Study): a randomized, double-blind, placebo-controlled trial. *Arch Neurol.* 2012;69(8):1044-1051.
15. Wesnes K, Myhr KM, Riise T, et al. Low vitamin D, but not tobacco use or high BMI, is associated with long-term disability progression in multiple sclerosis. *Mult Scler Relat Disord.* 2021;50:102801.
16. Connor Gorber S, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. *Nicotine Tob Res.* 2009;11(1):12-24.
17. Biochemical verification of tobacco use and cessation. *Nicotine Tob Res.* 2002;4(2):149-159.
18. Hedström AK, Bäärnhielm M, Olsson T, Alfredsson L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. *Neurology.* 2009;73(9):696-701.
19. Jiang W, St-Pierre S, Roy P, Morley BJ, Hao J, Simard AR. Infiltration of CCR2+ Ly6Chigh proinflammatory monocytes and neutrophils into the central nervous system is modulated by nicotinic acetylcholine receptors in a model of multiple sclerosis. *J Immunol.* 2016;196(5):2095-2108.
20. Friedman JI, Tang CY, de Haas HJ, et al. Brain imaging changes associated with risk factors for cardiovascular and cerebrovascular disease in asymptomatic patients. *JACC Cardiovasc Imaging.* 2014;7(10):1039-1053.
21. Ambrose JA, Barua RS. The pathophysiology of cigarette smoking and cardiovascular disease: an update. *J Am Coll Cardiol.* 2004;43(10):1731-1737.
22. Schmidt P, Gaser C, Arsic M, et al. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *NeuroImage.* 2012;59(4):3774-3783.
23. FMRIB Software Library. fmrib.ox.ac.uk/fsl.
24. Battaglini M, Jenkinson M, De Stefano N. Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Hum Brain Mapp.* 2012;33(9):2062-2071.
25. FreeSurfer. surfer.nmr.mgh.harvard.edu/.
26. Fischl B. FreeSurfer. *Neuroimage.* 2012;62(2):774-781.
27. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage.* 1999;9(2):179-194.
28. Lie IA, Kerklingh E, Wesnes K, et al. The effect of gadolinium-based contrast-agents on automated brain atrophy measurements by FreeSurfer in patients with multiple sclerosis. *Eur Radiol.* 2022;32(5):3576-3587.
29. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage.* 2006;31(3):968-980.
30. Hochmeister S, Grundtner R, Bauer J, et al. Dysferlin is a new marker for leaky brain blood vessels in multiple sclerosis. *J Neuropathol Exp Neurol.* 2006;65(9):855-865.
31. Absinta M, Sati P, Masuzzo F, et al. Association of chronic active multiple sclerosis lesions with disability in vivo. *JAMA Neurol.* 2019;76(12):1474-1483.
32. Dal-Bianco A, Grabner G, Kronnerwetter C, et al. Long-term evolution of multiple sclerosis iron rim lesions in 7 T MRI. *Brain.* 2021;144(3):833-847.
33. Elbejjani M, Auer R, Jacobs DR, et al. Cigarette smoking and gray matter brain volumes in middle age adults: the CARDIA Brain MRI sub-study. *Transl Psychiatry.* 2019;9(1):78.

34. Gallinat J, Meisenzahl E, Jacobsen LK, et al. Smoking and structural brain deficits: a volumetric MR investigation. *Eur J Neurosci.* 2006;24(6):1744-1750.
35. 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.
36. Bergsland N, Horakova D, Dwyer MG, 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.* 2012;33(8):1573-1578.
37. Henry RG, Shieh M, Amirbekian B, Chung S, Okuda DT, Pelletier D. Connecting white matter injury and thalamic atrophy in clinically isolated syndromes. *J Neurol Sci.* 2009;282(1-2):61-66.
38. Klaver R, De Vries HE, Schenk GJ, Geurts JJ. Grey matter damage in multiple sclerosis: a pathology perspective. *Prion.* 2013;7(1):66-75.
39. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. *Nat Rev Immunol.* 2015;15(9):545-558.
40. Cortese M, Munger KL, Martínez-Lapiscina EH, et al. Vitamin D, smoking, EBV, and long-term cognitive performance in MS. *Neurology.* 2020;94(18):e1950.
41. Karama S, Ducharme S, Corley J, et al. Cigarette smoking and thinning of the brain's cortex. *Mol Psychiatry.* 2015;20(6):778-785.
42. Manouchehrinia A, Tench CR, Maxted J, Bibani RH, Britton J, Constantinescu CS. Tobacco smoking and disability progression in multiple sclerosis: United Kingdom cohort study. *Brain.* 2013;136(pt 7):2298-2304.
43. Koch M, van Harten A, Uyttenboogaart M, De Keyser J. Cigarette smoking and progression in multiple sclerosis. *Neurology.* 2007;69(15):1515-1520.
44. Meyer-Moock S, Feng YS, Maeurer M, Dippel FW, Kohlmann T. Systematic literature review and validity evaluation of the expanded disability status scale (EDSS) and the multiple sclerosis functional composite (MSFC) in patients with multiple sclerosis. *BMC Neurol.* 2014;14:58.
45. Ozcan ME, Ince B, Bingöl A, et al. Association between smoking and cognitive impairment in multiple sclerosis. *Neuropsychiatr Dis Treat.* 2014;10:1715-1719.
46. Chiaravalloti ND, DeLuca J. Cognitive impairment in multiple sclerosis. *Lancet Neurol.* 2008;7(12):1139-1151.

The Effect of Smoking on Long-term Gray Matter Atrophy and Clinical Disability in Patients with Relapsing-Remitting Multiple Sclerosis

Ingrid Anne Lie, Kristin Wesnes, Silje S. Kvistad, et al.

Neurol Neuroimmunol Neuroinflamm 2022;9;
DOI 10.1212/NXI.0000000000200008

Supplemental data

Contents

eTable 1. Details on MRI acquisition per protocol

eTable 1. Details on MRI acquisition per protocol.										
Protocol (N of patients)	**1** (3)		**2** (13)		**3** (3)		**4** (1)		**5** (3)	
Scanner	Siemens Aera		Siemens Prisma		Siemens Skyra		Siemens Avanto		Siemens Skyra	
Field strength	1.5T		3T		3T		1.5T		3T	
Sequences	T1; MPRAGE	T2; FLAIR	T1; MPRAGE	T2; FLAIR	T1; MPRAGE	T2; FLAIR	T1; MPRAGE	T2; FLAIR	T1; MPRAGE	T2; FLAIR
TR (ms)	1940	5000	1800	5000	2300	5000	2060	5000	2300	5000
TE (ms)	2.69	335	2.28	386	2.32	387	3.10	340	2.32	387
TI (ms)	976	1800	900	1800	900	1800	1100	1800	900	1800
Flip angle (°)	8	120	8	120	8	120	15	120	8	120
Voxel size (mm)	1.00x0.98 x0.98	1.00x1.00x1.00	1.00x1.00x1.00	1.00x1.00x1.00	0.9x0.94 x0.94	0.9x0.45x0.45	1.00x0.49x0.49	1.00x0.51x0.51	0.90x0.94x0.94	0.90x0.45x0.45
Protocol (N of patients)	**6** (4)		**7** (3)		**8** (5)		**9** (12)		**10** (2)	
Scanner	Siemens Avanto		Siemens Aera		Philips Achieva		Philips Achieva		Siemens Prisma	
Field strength	1.5T		1.5T		1.5T		3T		3T	
Sequenses	T1; MPRAGE	T2; FLAIR	T1; MPRAGE	T2; FLAIR	T1; FFE	T2; FLAIR	T1; FFE	T2; FLAIR	T1; MPRAGE	T2; FLAIR
TR (ms)	2200	6000	2200	5000	7.6	4800	8.04	5000	1800	5000
TE (ms)	2.82	358	2.67	335	3.75	338	3.68	386	2.28	385
TI (ms)	900	2200	900	1800		1660		1650	900	1800
Flip angle (°)	8	120	8	120	8	90	8	90	8	120
Voxel size (mm)	1.00x0.98 x0.98	1.00x1.00x1.00	1.00x0.98 x0.98	1.00x1.00x1.00	1.00x0.98 x0.98	1.00x0.98x0.98	1.00x0.46x0.46	0.50x0.73x0.73	1.00x0.50 x0.50	1.00x0.50x0.50
Protocol (N of patients)	**11** (6)		**12** (1)		**13** (2)		**14** (8)		**15** (3)	
Scanner	Philips Ingenia		Toshiba MRT200SP3		Philips Ingenia		Siemens Prisma		Philips Achieva	

Field strength	1.5T		1.5T		3T		3T		1.5T	
Sequenses	T1; FFE	T2; FLAIR	T1; FFE	T2; FLAIR	T1; FFE	T2; FLAIR	T1; MPRAGE	T2; FLAIR	T1; FFE	T2; FLAIR
TR (ms)	25	4800	13.5	1160	11.11	4800	1800	5000	7.1	4800
TE (ms)	9.21	367	5.50	105	6.29	296	2.26	387	2.2	307
TI (ms)		1660		2300		1650	906	1800		1660
Flip angle (°)	30	90	20	90	8	90	8	120	8	90
Voxel size (mm)	1.00x0.94x0.94	0.50x0.74x0.74	0.39x0.65x0.39	0.47x4.0x0.47	0.90x0.67x0.67	0.56x0.98x0.98	1.00x1.000x1.00	1.00x1.000x1.00	1.00x1.00 x1.00	0.76x1.00x1.00

Abbreviations: N=number; T=Tesla; MPRAGE=magnetization-prepared rapid gradient-echo; FLAIR=fluid-attenuated inversion recovery; TR=repetition time; TE=echo time; TI=inversion time; ms=millisecond; mm=millimeter; FFE=fast field echo.

eTable 2 Smoking habits of smokers defined by patient self-reporting (N = 48)				
Current smoker	*Yes:* 27 (56.3%)		*No:* 21 (43.8%)	
Smoking 10 years ago	*Yes:* 47 (97.9%)		*No:* 1 (2.1%)	
Smoking frequency 10 years ago	*1-3 days / month* 2 (4.3%)	*1-2 days / week* 3 (6.4%)	*3-6 days / week* 2 (4.3%)	*Daily* 40 (85.1%)
Daily number of cigarettes 10 years ago	*1-4 / day* 4 (8.5%)	*5-10 / day* 17 (36.2%)	*11-20 / day* 23 (48.9%)	*>20 /day* 3 (6.4%)

Abbreviations: N=number, SD=standard deviation

eTable 3 MRI and clinical measures in smokers and non-smokers (defined by patient self-reporting) at the 10-year follow-up visit and change in clinical measures between month 24 and the 10-year follow-up visit.

MRI/clinical measure	*Non-smokers (mean, SD)*	*Smokers (mean, SD)*	*Mean difference*	*95% conf. interval*	*p-value*
Total GM volume (mL)	643.48 (57.41)	620.84 (45.45)	22.64	-2.26, 47.53	0.074
Total WM volume (mL)	469.65 (55.90)	433.78 (39.83)	35.87	12.68, 59.07	**0.003**
Total deep GM volume (mL)	58.09 (4.88)	54.37 (5.33)	3.73	1.22, 6.23	**0.004**
Thalamus volume (mL)	8.09 (0.98)	7.66 (1.01)	0.42	-0.06, 0.91	0.086
Mean Cth (mm)	2.55 (0.13)	2.53 (0.13)	0.02	-0.04, 0.09	0.424
T2 Lesion volume (mL)[1]	3.06 (3.28)	6.61 (8.71)			**0.021**
T2 Lesion count	19.56 (7.73)	21.70 (9.70)	-2.14	-6.34, 2.06	0.314
EDSS[1]	2.5 (1.0)	2.5 (2.0)			0.419
Change in EDSS[1]	0.50 (1.50)	0.50 (1.13)			0.474
T25FW[1]	3.88 (1.13)	4.48 (1.32)			**0.005**
Change in T25FW[1]	-0.13 (1.11)	0.43 (1.33)			**0.021**
D9-HPT[1]	20.13 (6.18)	21.15 (5.59)			0.351
Change in D9-HPT[1]	3.10 (2.72)	2.82 (5.90)			0.700
ND9-HPT[1]	20.71 (3.92)	22.40 (8.35)			0.100
Change in ND-9HPT[1]	1.80 (3.78)	4.09 (6.36)			0.069
PASAT	49.30 (7.84)	45.28 (10.91)	4.02	-0.41, 8.45	0.075
Change in PASAT	-3.47 (6.68)	-7.39 (7.03)	3.92	0.67, 7.17	**0.019**

[1]Difference analyzed by Mann-Whitney U test, median and interquartile range reported.
Abbreviations: SD=standard deviation; conf=confidence; GM=gray matter; mL=milliliter; WM=white matter; Cth=cortical thickness; mm=millimeter; EDSS=expanded disability status scale; T25FW=timed 25-foot walk; D9-HPT=dominant hand 9-hole peg test; ND9-HPT=non-dominant hand 9-hole peg test; PASAT=paced auditory serial addition test.

eTable 4 The effect of smoking status (defined by patient self-reporting) on MRI and clinical measures at the 10-year follow-up visit and change in clinical measures between month 24 and the 10-year follow-up visit.

MRI/clinical measure	*N*	*Beta*	*St. Error*	*95% confidence interval*	*p-value*
Total GM volume (mL)	66	-11.002	9.700	-28.825, 6.820	0.263
Total WM volume (mL)	66	-24.918	10.563	-44.326, -5.510	**0.023**
Total deep GM volume (mL)	66	-2.349	1.163	-4.485, -0.213	**0.049**
Thalamus volume (mL)	66	-0.305	0.227	-0.723, 0.112	0.186
Mean Cth (mm)	66	-0.020	0.025	-0.067, 0.027	0.443
LogT2 Lesion volume (mL)[1]	63	0.208	0.087	0.047, 0.368	**0.022**
T2 Lesion count	63	1.969	2.390	-2.448, 6.385	0.415
EDSS	66	0.195	0.288	-0.338, 0.727	0.502
Change in EDSS	64	-0.064	0.312	-0.641, 0.513	0.838
LogT25FW[1]	65	0.050	0.024	0.007, 0.094	**0.039**
LogChange in T25FW[1]	63	0.040	0.020	0.002, 0.078	0.056
LogD9-HPT[1]	63	0.014	0.024	-0.031, 0.058	0.567
LogChange in D9-HPT[1]	61	0.033	0.051	-0.061, 0.126	0.525
LogND9-HPT[1]	63	0.029	0.021	-0.010, 0.069	0.172
LogChange in ND9-HPT[1]	61	0.120	0.062	0.005, 0.234	0.062
PASAT	63	-2.970	2.205	-7.045, 1.105	0.185
Change in PASAT	61	-3.575	1.577	-6.482, -0.667	**0.029**

[1]Dependent variable log transformed due to non-normality (log-linear transformation).

Abbreviations: N=number; St= standard; GM=gray matter; mL=milliliter; WM=white matter; Cth=cortical thickness; mm=millimeter; EDSS=expanded disability status scale; T25FW=timed 25-foot walk; D9-HPT=dominant hand 9-hole peg test; ND9-HPT=non-dominant hand 9-hole peg test; PASAT=paced auditory serial addition test.

eTable 5 Dose-effect relationship between mean cotinine level in smokers and MRI and clinical measures at the 10-year follow-up visit and change in clinical measures between month 24 and the 10-year follow-up visit.

MRI/clinical measure	*N*	*Beta*	*St. Error*	*95% confidence interval*	*p-value*
Total GM volume (mL)	36	-0.023	0.017	-0.053, 0.007	0.211
Total WM volume (mL)	36	-0.006	0.018	-0.038, 0.026	0.741
Total deep GM volume (mL)	36	-0.003	0.002	-0.006, 0.001	0.238
Thalamus volume (mL)	36	-0.001	$3.994*10^{-4}$	-0.001, $3.147*10^{-5}$	0.119
Mean Cth (mm)	36	$-6.990*10^{-5}$	$5.019*10^{-5}$	$-1.582*10^{-4}$, $1.839*10^{-5}$	0.182
LogT2 Lesion volume (mL)[1]	34	$1.593*10^{-4}$	$1.622*10^{-4}$	$-1.241*10^{-4}$, $4.426*10^{-4}$	0.341
T2 Lesion count	34	0.008	0.005	$3.991*10^{-4}$, 0.016	0.085
EDSS	36	$1.147*10^{-4}$	$3.907*10^{-4}$	-0.001, 0.001	0.773
Change in EDSS	35	$-2.540*10^{-4}$	$3.897*10^{-4}$	-0.001, $4.287*10^{-4}$	0.523
LogT25FW[1]	35	$-3.500*10^{-6}$	$3.851*10^{-5}$	$-7.102*10^{-5}$, $6.401*10^{-5}$	0.929
LogChange in T25FW[1]	34	$-2.270*10^{-5}$	$2.141*10^{-5}$	$-6.007*10^{-5}$, $1.472*10^{-5}$	0.305
LogD9-HPT[1]	33	$-4.750*10^{-5}$	$3.925*10^{-5}$	$-1.158*10^{-4}$, $2.079*10^{-5}$	0.245
LogChange in D9-HPT[1]	32	$-1.028*10^{-4}$	$6.805*10^{-5}$	$-2.207*10^{-4}$, $1.509*10^{-5}$	0.153
LogND9-HPT[1]	33	$-1.990*10^{-5}$	$3.789*10^{-5}$	$-8.586*10^{-5}$, $4.597*10^{-5}$	0.606
LogChange in ND9-HPT[1]	32	$-1.032*10^{-4}$	$9.993*10^{-5}$	$-2.763*10^{-4}$, $6.988*10^{-5}$	0.319
PASAT	34	-0.007	0.004	-0.014, $1.642*10^{-4}$	0.106
Change in PASAT	33	-0.001	0.003	-0.006, 0.004	0.771

[1]Dependent variable log transformed due to non-normality (log-linear transformation).

Abbreviations: eTIV=estimated total intracranial volume; N=number; St= standard; GM=gray matter; mL=milliliter; WM=white matter; Cth=cortical thickness; mm=millimeter; EDSS=expanded disability status scale; T25FW=timed 25-foot walk; D9-HPT=dominant hand 9-hole peg test; ND9-HPT=non-dominant hand 9-hole peg test; PASAT=paced auditory serial addition test.

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230855478 (print)
9788230860274 (PDF)