

Improving the clinico-radiological association in neurological diseases

Inaugural dissertation

to
be awarded the degree of Dr. sc. med.
presented at
the Faculty of Medicine
of the University of Basel

by
Anna B. Altermatt
from Nunningen, Switzerland

Basel, 2018

Approved by the Faculty of Medicine
on application of

Prof. Philippe Cattin, University of Basel - *Primary advisor*

Prof. Roland Wiest, University of Bern - *External advisor*

Dr. Jens Würfel, University of Basel - *Secondary advisor*

Dr. Laura Gaetano, University of Basel - *Tertiary advisor*

Basel, June 26th, 2018

Prof. Thomas Gasser - *Dean*

Acknowledgments

Zuerst gebührt mein Dank Herrn Prof. Philippe Cattin und Herrn Dr. Jens Würfel, die meine Arbeit während den vergangenen drei Jahren betreut haben. Die Möglichkeit mein Doktorat an einer Schnittstelle der Forschung der medizinischen Fakultät der Universität Basel und der Medical Image Analysis Center (MIAC) AG zu realisieren, empfand ich als sehr inspirierend und dafür möchte ich mich gerne ganz herzlich bedanken.

Ein speziell grosses Dankeschön geht an meine direkte Betreuerin innerhalb unseres Forschungsteams, Frau Dr. Laura Gaetano, die meine Arbeit richtungsweisend und mit viel Engagement unterstützt hat. Laura hat es verstanden mir in allen Phasen der Projekte motivierend zur Seite zu stehen und für die zahlreichen Inputs, den wertvollen Austausch und ihre wohlwollende Betreuung möchte ich mich herzlich bedanken. Dann möchte ich mich gerne bei Herrn Dr. Francesco Santini und Frau Dr. Xenia Deligianni bedanken, die mir vor allem mit ihrer Physik Expertise zur Seite standen und mich jederzeit äusserst kompetent und bereitwillig unterstützt haben. Ein weiterer Dank geht an meine geschätzten Teamkolleginnen und Kollegen. Ein spezieller Dank gilt hier Herrn Dr. Stefano Magon, der mich vor allem bei den statistischen Analysen tatkräftig unterstützt hat. Ein weiterer Dank geht an meinen Kollegen, Haris Tsagkas, als meinen Ansprechpartner in medizinischen Fragen, als Mitstreiter und Freund. Des Weiteren bedanke ich mich bei allen verbleibenden Kolleginnen und Kollegen und allen Mitarbeiterinnen und Mitarbeitern der MIAC AG, die allesamt dazu beigetragen haben, dass mir die drei Doktoratsjahre als positiv und erfreulich in Erinnerung bleiben werden.

Grosse Dinge gedeihen und gelingen leichter mit kräftigen Wurzeln und in der richtigen Umgebung. Deshalb möchte ich mich an dieser Stelle auch ganz herzlich bei meiner Familie und Verwandtschaft bedanken. Der allergrösste Dank gebührt an dieser Stelle meinem Grossvater, Wilhelm Grolimund, für die kräftigsten Wurzeln und die offenen Arme in schwierigen Zeiten. Des Weiteren danke ich meinen lieben Freundinnen und Freunden, insbesondere Sibylle Meier, die mir eine wahrhaftige Freundin in allen Lebenslagen ist. Tief dankbar bin ich schlussendlich meinem Partner, Matthias Sutter, für das ermutigendste und breiteste Lachen in meinem Leben.

Summary

Despite the key role of magnetic resonance imaging (MRI) in the diagnosis and monitoring of multiple sclerosis (MS) and cerebral small vessel disease (SVD), the association between clinical and radiological disease manifestations is often only moderate, limiting the use of MRI-derived markers in the clinical routine or as endpoints in clinical trials. In the projects conducted as part of this thesis, we addressed this clinico-radiological gap using two different approaches.

Lesion-symptom association: In two voxel-based lesion-symptom mapping studies, we aimed at strengthening *lesion-symptom* associations by identifying strategic lesion locations. Lesion mapping was performed in two large cohorts: a dataset of 2348 relapsing-remitting MS patients, and a population-based cohort of 1017 elderly subjects. T2-weighted lesion masks were anatomically aligned and a voxel-based statistical approach to relate lesion location to different clinical rating scales was implemented. In the MS lesion mapping, significant associations between white matter (WM) lesion location and several clinical scores were found in periventricular areas. Such lesion clusters appear to be associated with impairment of different physical and cognitive abilities, probably because they affect commissural and long projection fibers. In the SVD lesion mapping, the same WM fibers and the caudate nucleus were identified to significantly relate to the subjects' cerebrovascular risk profiles, while no other locations were found to be associated with cognitive impairment.

Atrophy-symptom association: With the construction of an anatomical physical phantom, we aimed at addressing reliability and robustness of *atrophy-symptom* associations through the provision of a “ground truth” for atrophy quantification. The built phantom prototype is composed of agar gels doped with MRI and computed tomography (CT) contrast agents, which realistically mimic T1 relaxation times of WM and grey matter (GM) and showing distinguishable attenuation coefficients using CT. Moreover, due to the design of anatomically simulated molds, both WM and GM are characterized by shapes comparable to the human counterpart. In a proof-of-principle study, the designed phantom was used to validate automatic brain tissue quantification by two popular software tools, where “ground truth” volumes were derived from high-resolution CT scans. In general, results from the same software yielded reliable and robust results across scans, while results across software were highly variable reaching volume differences of up to 8%.

Table of Contents

Acknowledgments	iii
Summary	v
1 Introduction	1
1.1 Motivation	1
1.2 Contribution	1
1.3 Outline	2
2 Medical background	5
2.1 Multiple sclerosis	5
2.1.1 Disease overview	5
2.1.2 Role of magnetic resonance imaging	6
2.2 Cerebral small vessel disease	8
2.2.1 Disease overview	8
2.2.2 Role of magnetic resonance imaging	9
3 Relevance of lesion location and lesion-symptom mapping	11
3.1 Clinico-radiological association	11
3.2 Relevance of lesion location and voxel-based lesion-symptom mapping	12
3.3 Voxel-based lesion-symptom mapping in multiple sclerosis	13
3.4 Voxel-based lesion-symptom mapping in cerebral small vessel disease	15
4 Two large cohort voxel-based lesion-symptom mapping studies	17
5 Relevance of segmentation validation in brain atrophy measurements	45
5.1 Clinico-radiological association	45
5.2 Relevance of segmentation validation in brain atrophy measurements	47
5.3 Anthropomorphic physical phantoms	48
6 An anthropomorphic physical phantom for segmentation validation	51
7 Discussion and Conclusions	79

1 Introduction

1.1 Motivation

Magnetic resonance imaging (MRI) is a key tool in supporting the diagnosis and monitoring of many neurological diseases, including multiple sclerosis (MS) and cerebral small vessel disease (SVD) [1, 2]. In both diseases, local white matter (WM) lesions are the classic hallmark pathologies. Damage also occurs in grey matter (GM) as well as diffusely in normal appearing WM, and may lead to regional or total brain volume loss called atrophy [3, 4]. Despite the central role of MRI in assessing these pathologies, the association between clinical and radiological disease manifestations is often only moderate [5–7]. This limits the use of MRI-derived markers in the clinical routine or as endpoints in clinical trials. In our projects, we aimed at bridging this clinico-radiological gap using two different approaches.

Lesion-symptom association: By applying voxel-based lesion-symptom mapping to two large cohorts of MS and SVD patients, we aimed at strengthening the *lesion-symptom* associations by identifying strategic lesion locations.

Atrophy-symptom association: With the construction of an anatomical physical phantom we aimed at providing a "ground truth" for the validation of atrophy measurements, which will allow to address the robustness and reliability of *atrophy-symptom* associations.

1.2 Contribution

Lesion-symptom association: Voxel-based lesion-symptom mapping was performed in two large cohorts: a dataset of 2348 relapsing-remitting MS patients, and a population-based cohort of 1017 elderly subjects showing lesions characteristic of cerebral SVD. For both studies, an automatized registration pipeline for alignment of the T2-weighted lesion masks was implemented. In a second step, a voxel-based statistical approach to relate lesion location to different clinical measures was encoded.

Previous voxel-based lesion-symptom mapping studies have already demonstrated their validity and superiority over classical lesion-symptom correlation studies in both MS [8–14] and SVD [15–19]. However, results across studies were partially inconsistent, some analyses produced non-specific results, lesion-symptom associations were sometimes difficult to interpret, and some studies even failed to demonstrate any relevance of lesion location for

some clinical rating scales. This might partly relate to methodological issues linked to the stereotaxic alignment of lesion masks and the conventional parametric statistical analyses employed in previous studies. Indeed, in several previous studies, lesion masks were aligned to a standard brain template using linear transformations only, potentially introducing bias regarding the exact anatomical locations [20]. Moreover, the binary nature of the lesion maps was not respected, e.g. if conventional parametric statistical models were applied [14]. Finally, even in study cohorts of up to several hundred patients, voxel-based lesion mapping approaches may not allow for sufficient statistical power in regions with low lesion coverage [7]. In this regard, large cohorts including up to thousand or more patients depict a more promising framework to study lesion-symptom associations. For these reasons, we here propose two optimized preprocessing and data analysis pipelines allowing for robust automatized lesion-symptom mapping in big sample input data.

Atrophy-symptom association: We here suggest the design of an anthropomorphic physical phantom mimicking brain WM and GM, which, after its construction, is used in a proof-of-principle study to validate two commonly-used software tools for the quantification of atrophy.

In previous validation studies, it was shown that the choice of the algorithm for automatized atrophy quantification had the largest impact on variability of computed WM and GM volumes. Since the discrepancies between results reach the same order of magnitude as volume changes observed in disease, these software effects were shown to limit the usability of these methods for following volume changes in individual patients over time [21, 22]. To date, validation of these methods for the quantification of atrophy has never been done using a physical phantom, which realistically mimics T1 relaxation times and structure of the main brain compartments. A number of multipurpose physical phantoms mimicking brain T1 relaxation times and showing increased anatomical similarity to the real counterpart have been developed by research groups or academic centers [23–28]. These existing physical phantoms were hampered by a few limitations – most importantly the substantial simplification of brain anatomy or the presence of physical (e.g. 3D-printed) walls separating the phantom compartments. In this regard, due to the increased anatomical similarity and the absence of in-built compartments, the here proposed phantom allows to more realistically mimic atrophy quantification compared to existing phantoms. Moreover, following the validation of two software, we are able to identify the algorithm that provides more robust and reliable results for future studies.

1.3 Outline

In Chapter 2, the reader is made familiar with the medical background of MS and cerebral SVD, emphasizing on the role of MRI in these two neurological diseases.

Chapters 3 and 4 focus on the conducted voxel-based lesion-symptom mapping studies. Chapter 3 starts by describing the mismatch between disease severity and conventional lesion-derived metrics, discusses the role of strategic lesion location in strengthening the associations between clinical and radiological disease manifestations, introduces the concept

behind voxel-based lesion-symptom mapping, and references to prior work that has been done in this context. Chapter 4 describes the two publications on the voxel-based lesion-symptom mapping studies, which were conducted as part of this thesis.

Chapters 5 and 6 describe the construction and application of the designed anatomical physical phantom used for the validation of brain atrophy measurements. Chapter 5 gives an introduction on the role of atrophic disease manifestations in explaining clinical disability in neurological disease, describes different established methods used for the validation of brain tissue quantification, and refers to existing physical phantoms mimicking shape and MR relaxation times of brain WM and GM compartments. Chapter 6 comprises the second part of the thesis on the construction of an anatomical phantom used for the validation of brain atrophy measurements, which resulted in two publications: a first manuscript on the phantom construction, and a proof-of-principle study where the designed phantom is used to validate the quantification of WM and GM atrophy by two commonly-used software tools.

The thesis is completed by a discussion and conclusion in Chapter 7.

2 Medical background

This chapter provides a brief introduction on some important aspects of the two neurological diseases MS (Section 2.1) and cerebral SVD (Section 2.2), including epidemiology, etiology, symptoms and subtypes, pathogenesis, and diagnosis. Most importantly, the radiological manifestations of both diseases are discussed.

2.1 Multiple sclerosis

2.1.1 Disease overview

MS is a chronic, immune-mediated, demyelination disorder of the central nervous system [29]. An early detailed description of MS dates back to the 19th century and the French neurologist Jean-Martin Charcot. Indeed, Charcot’s description of MS (described as “la sclérose en plaques”) in 1868, accompanied by the first illustrations on the expansion of the so-called plaques from the ventricles into the cerebral hemispheres, provided the earliest insight into the pathology of MS [30].

Epidemiology: MS is the most common cause of nontraumatic disability in young adults. The incidence of MS is low during childhood and increases after the age of 18, reaching a maximum between the age of 20-40 [31]. MS affects women more than men (sex ratio 2.5:1) and the prevalence varies by geographic area. Across Europe, the prevalence rate of MS is about 83 in 10’000 individuals [32].

Etiology: The etiology of MS is largely unknown, however, epidemiological data indicate that both environmental (e.g. geographic latitude, tobacco exposure, obesity, and viral infections) and genetic factors (e.g. female gender, genes mainly associated with immune-pathway) play a key role in the disease development [29].

Symptoms and subtypes: The disease can present with changes in sensation, mobility, balance, vegetative functions, vision, and cognition [33]. It is classified as either relapsing-remitting or primary progressive based on the initial disease course. Relapsing-remitting MS is more common, affecting 85–90% of patients with MS, and is characterized by relapses followed by periods of remission of symptoms. Primary progressive MS, affecting 10–15% of patients, is characterized by a slowly progressive increase in neurological disability over time. Primary progressive MS typically presents at an older age than does relapsing-remitting MS. People with relapsing-remitting MS may develop a progressive course with time (secondary progressive MS) with a gradual increase in disability in this later disease phase [34].

Pathogenesis: The pathologic hallmark of MS is multiple focal areas of myelin damage and destruction within the central nervous system called plaques or lesions. Damage can occur throughout the central nervous system and is most easily recognized in the WM as focal areas of demyelination, inflammation, and glial reaction. MS pathology results from a complex and dynamic interplay between innate and adaptive immune system, glia (myelin-making oligodendrocytes, microglia, and astrocytes), and neurons. The traditional view is that myelin-specific autoreactive lymphocytes (mainly T helper cells), which are primed outside the central nervous system, cross the blood-brain barrier, and cause the formation of inflammatory demyelinating lesions [29]. A recent hypothesis, however, suggests that MS is an immunological convolution between a primary degenerative disorder and an aberrant immune response [35]. Besides focal lesions within the central nervous system, diffuse WM and GM damage and atrophy are observed. Atrophy refers to tissue loss through pathologic neurodegenerative processes. Volume loss can also arise from resolution of inflammatory edema and other pathological and physiological reductions in the water content of brain tissues [3].

Diagnosis: MRI has become a critically important tool in the diagnosis of MS, which is based on the demonstration of lesion dissemination in space (i.e. lesions at different MS-typical locations within central nervous system) and time (i.e. evidence of at least a second episode of disease activity). Diagnostic criteria for MS combining clinical, imaging, and laboratory evidence have evolved over time, with the most recent being described in the 2017 revisions of the McDonald criteria [1].

2.1.2 Role of magnetic resonance imaging

Since its innovation and introduction in the clinical practice in the early 1980s, MRI has quickly been adopted as an essential tool in supporting the diagnosis, longitudinal monitoring, evaluation of therapeutic response, and scientific investigations in MS. Although MS is a disease that predominantly affects the WM, different pathologies can be detected throughout the central nervous system. A number of MRI techniques are routinely used to identify or quantify MS pathology.

White matter lesions: The hallmark lesions in MS are within the WM of the brain (Figure 1). Lesions are typically round and are varying in size from a few millimeters to some centimeters. MS lesions mainly occur within periventricular regions, the corpus callosum, the juxtacortical WM-GM junction and the infratentorial brain regions (i.e. brain stem, cerebellar peduncles and cerebellar WM in the posterior fossa) and spinal cord [1, 36]. The classical sequences that allow to visualize MS plaques *in vivo* are T2-weighted imaging techniques, where lesions appear as hyperintense. Lesions detected with T2-weighted imaging have different pathophysiological correlates (e.g. demyelination, remyelination, inflammation, edema, Wallerian degeneration, axonal loss). The most common T2-weighted sequences are fast spin echo (FSE) based techniques and fluid attenuation inversion recovery (FLAIR). In the latter acquisition technique, an additional inversion recovery pulse is used to suppress the cerebrospinal fluid (CSF) signal allowing for an improved detection of

cerebral hemispheric lesions owing to an increased tissue contrast [37].

A subset of T2-weighted lesions appear hypointense in T1-weighted imaging. Indeed, approximately 30% of T1-weighted hypointensities, the so-called "black-holes" will persist and are associated with more severe tissue loss or axonal injury [37].

The contrast agent Gadolinium is used to detect blood brain barrier breakdown, an indirect sign of acute focal disease activity [37].

Cortical lesions: The detection of cortical lesions is typically done using T2-weighted imaging techniques, including FLAIR or double inversion recovery (DIR) (Figure 1). At standard field strengths, large part of cortical lesions remain undetected, especially if their location is purely intracortical. Studies have shown improved detection and localization of cortical lesions by using higher field strengths up to 7 Tesla [38].

Optic nerve lesions: The optic nerve is frequently involved in demyelinating disease. Optic neuritis is present in up to 50% of patients with MS and is frequently the presenting sign [37]. Fat-saturated T2-weighted imaging and contrast-enhanced T1-weighted imaging are sensitive methods to detect optic nerve lesions, which typically occur in the retrobulbar nerve segments (Figure 1) [39].

Spinal cord lesions: The spinal cord is also frequently involved in MS and, for most patients, both spinal cord and the brain are affected. Most spinal lesions are localized in the cervical rather than the thoracic cord. At MR imaging, spinal lesions show increased T2-weighted signal intensity and frequently also gadolinium enhancement (Figure 1) [40].

Brain atrophy: Brain atrophy, which is usually quantified on T1-weighted images, is another hallmark of MS (Figure 1). The rate of whole-brain atrophy in MS is 0.5 to 1% per year. It is not confined to specific structures but occurs throughout WM and GM and also the spinal cord [3].

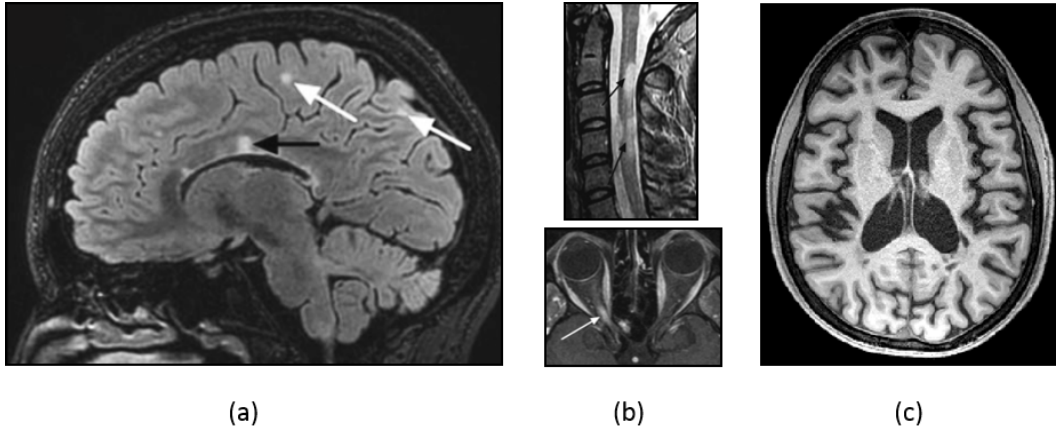


Fig. 1: Lesions characteristic of MS. (a) Sagittal image with WM lesions (black arrow) and juxtacortical lesions (white arrows) on T2-weighted FLAIR; (b) Other common lesion locations in MS include spinal cord (top image, black arrows, sagittal T2-weighted sequence) and optic nerve lesions (bottom image, white arrows, gadolinium-enhanced T1-weighted MRI); (c) Brain atrophy typically manifests through brain parenchymal shrinkage and is usually quantified in T1-weighted MRI (images taken from [37] and from MS Rehab study, University Hospital of Basel).

2.2 Cerebral small vessel disease

2.2.1 Disease overview

The term cerebral SVD is used with various meanings in different contexts, to describe a range of neuroimaging and pathological findings, as well as associated clinical and cognitive features or syndromes. However, in its most basic form, the term encompasses a range of pathological processes affecting the microvasculature of the brain and is one of the most common causes of cerebrovascular disease including stroke and vascular/mixed dementia [41].

Epidemiology: Studies have shown that SVD is a leading cause of cognitive decline in the elderly, accounts for 20–30% of all strokes, and is also thought to account for about 45% of dementia cases [2, 42]. SVD is more common in elderly people and seen more in women compared to men. It often progresses subclinically for many years until the individual presents with first symptoms. Since cerebral SVD is not a standardized concept yet but relates to a range of neuroimaging and pathological findings, the prevalence of cerebral SVD is not well established in the general population [41].

Etiology: The majority of SVD is sporadic and seems to be driven by a complex mixture of genetic and cardiovascular risk factors. Especially, arteriosclerosis is strongly associated with aging, diabetes, and hypertension. Such risk factors are likely to worsen disease progression via deleterious effects on both the structure and functioning of cerebral blood vessels [2]. A number of hereditary forms of cerebral SVD have also been identified [2, 43].

Symptoms and subtypes: Since the vasculature can be affected throughout the brain, the disease presents with a variety of symptoms. Besides cognitive impairment, other clinical characteristics mainly associated with SVD are gait, mood and behavioral, and urinary disturbances [2]. Different types of SVD are differentiated, among which arteriolosclerosis (i.e. type 1 SVD) and sporadic and hereditary cerebral amyloid angiopathy (i.e. type 2 SVD) are the most prevalent forms [2].

Pathogenesis: Arteriosclerosis and cerebral amyloid angiopathy are the most prevalent forms of SVD and, therefore, only the pathological characteristics of these diseases are described here. Arteriosclerosis, or type 1 SVD, is also known as age-related and vascular risk-factor-related SVD. From a pathological point of view, type 1 SVD is mainly characterized by loss of smooth muscle cells from the tunica media, deposits of fibro-hyaline material, narrowing of the vessel lumen, or thickening of the vessel wall [2]. Cerebral amyloid angiopathy, or type 2 SVD, is characterized by the progressive accumulation of amyloid protein in the walls of small-to-medium-sized arteries and arterioles. Importantly, cerebral amyloid angiopathy is a pathological hallmark of Alzheimer’s disease, in which it is almost invariably seen [2]. The mechanisms that link SVD with brain damage are heterogeneous and not completely understood. In general, the described pathological changes in all types of SVD can lead to both hemorrhage and ischemia [2]. An increased rate in brain volume loss is also associated with SVD. The mechanism underlying atrophy in cerebral SVD is not fully understood yet. In general, axonal loss, resolution of inflammation and edema, gliosis, demyelination, dehydration, and normal aging can result in a decrease in brain volume. No direct pathological studies have been performed yet looking at the histology underlying atrophy in SVD [6].

Diagnosis: The diagnosis is based on the underlying pathological changes of small vessels (e.g. including arteries, arterioles, capillaries) of the brain. Clinical findings include WM hyperintensities, small infarctions or hemorrhages in WM or deep GM, and brain atrophy, which are typically detected using MRI or CT [41]. One of the difficulties in diagnosing cerebral SVD is that these markers are not specific for SVD alone. Therefore, clinicians rely on the presence of a combination of these features for proper diagnosis of the disease [44]. Yet, the definition of SVD is not uniform and so is its diagnosis. Neuroimaging standards only begin to be established [4].

2.2.2 Role of magnetic resonance imaging

The effect of SVD on the brain parenchyma is usually derived from MRI or CT, and these changes are considered the hallmarks of the disease [41]. As mentioned earlier, the consequences of cerebral SVD on the brain tissue are heterogeneous, including ischemic and haemorrhagic manifestations [2]. A number of MRI techniques are routinely used to identify or quantify these pathologies (Figure 2).

White matter lesions: The ischemic WM lesions on MRI are seen as more or less confluent areas that are bilaterally and symmetrically sited in the brain parenchyma and appear hyperintense on T2-weighted and FLAIR images (Figure 2) [2]. Typical lesion locations

include the periventricular WM of the cerebral hemispheres, basal ganglia, pons, brainstem, and cerebellum [41]. Lesion location is thought to differ based on SVD etiology [45]. WM lesions have a variety of pathological correlates depending on the severity of ischemic tissue damage: myelin pallor, gliosis, axonal loss, complete nerve fiber destruction, and, in the worst cases, blood-brain barrier disruption and loss of endothelium [41].

Lacunar infarcts: Acutely, these infarcts appear as hyperintense on diffusion-weighted imaging (DWI), and within hours to days on T2-weighted imaging or FLAIR sequences. Chronic infarcts or lacunes appear hypointense on T1-weighted images and FLAIR [41]. They are typically found in the basal ganglia, internal capsule, thalamus, and pons (Figure 2) [2].

Cerebral microbleeds: This phenotype of SVD refers to small hemorrhages of 2-10 mm in diameter seen by MRI. The T2*-weighted gradient echo sequence, and the newer susceptibility-weighted imaging (SWI), are sensitive methods in detecting these (Figure 2) [41]. Microbleeds can occur throughout the brain, and their location seems to differ based on SVD etiology [45].

Atrophy: An alternative, promising marker that may be computed from conventional T1-weighted MRI sequences is whole or regional brain volume. A previous longitudinal follow-up study has demonstrated that the rate of atrophy in patients with SVD is approximately 1% per year and twice the one found in age-matched control subjects. However, to date, defined cut-offs regarding whether atrophy is related to normal aging or SVD are missing [6].

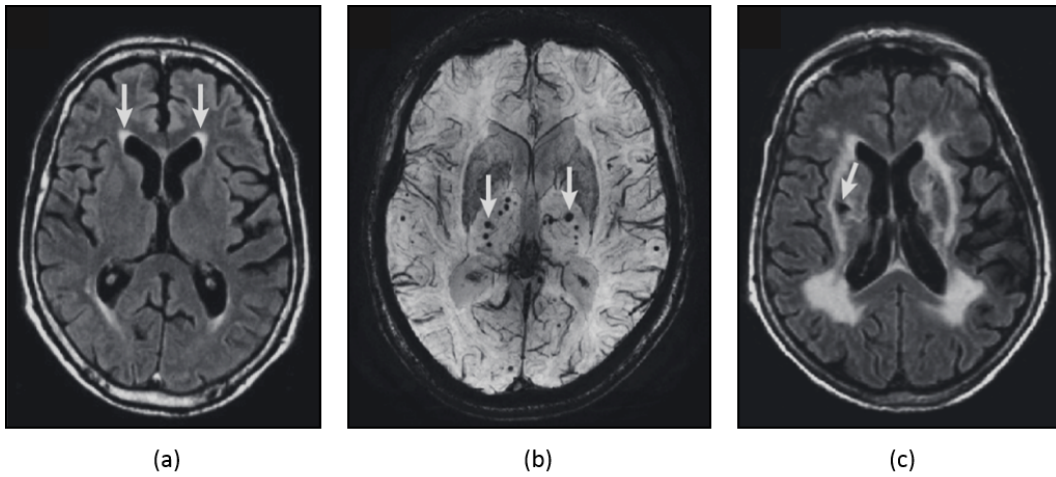


Fig. 2: Lesions characteristic of cerebral SVD. (a) Axial image with periventricular WM hyperintense lesions on FLAIR MRI; (b) Axial T2*-weighted image showing microbleeds as rounded hypointensities in the basal ganglia (white arrows) and the cortex; (c) An axial FLAIR image showing a lacunar infarct in the deep WM and GM (white arrow; images taken from [36]).

3 Relevance of lesion location and lesion-symptom mapping

As described in Chapter 2, MRI is a key tool in the diagnosis and monitoring of both MS and cerebral SVD, based on its ability to show lesions in different compartments of the central nervous system. However, the number and volume of (especially brain WM) lesions explain only a small fraction of the diversity of symptoms. In Section 3.1, different factors contributing to this clinico-radiological mismatch are discussed. Section 3.2 explains how lesion location may be taken into account in voxel-based lesion-symptom mapping approaches. Then, existing lesion-symptom mapping approaches in MS (Section 3.3) and SVD (Section 3.4) are presented.

3.1 Clinico-radiological association

Clinico-radiological association: The clinical course of MS and SVD may be diverse and unpredictable: Some people develop rapidly evolving impairment, whereas others accrue little or no detectable neurological disability over years or decades [5, 7]. In both diseases, MRI plays a key role in diagnosis and monitoring [1, 4]. However, clinico-radiological associations are often weak on a single subject level when regarding the total WM lesion burden. Indeed, this is illustrated by the frequent observation that some individuals with severe disease manifestation on MR images show little to no symptoms, while others with only a moderate visible lesion burden can be affected more severely [5, 7]. In MS, this observation has occupied researchers since decades and the mismatch between clinically detectable symptoms and the presence or amount of neuroradiological manifestations on MRI was termed the “clinico-radiological paradox” in the early 2000s [46]. In SVD, since clinically silent WM lesions can frequently be found in the elderly population, the clinical significance of WM lesions was even doubted until several case reports and small cohort studies repeatedly described an association between these WM lesions and cognitive impairment [7].

Factors contributing to moderate association: Today, several themes and issues are thought to underlie this clinico-radiological mismatch including the relevance of strategic lesion location, the low histopathological specificity of some neuroradiological markers, the underestimation of lesions with conventional MRI (such as cortical and subcortical GM lesions or damage to normal-appearing WM), the relevance of an integrated network underlying brain function, or cortical adaptive changes [5, 46].

3.2 Relevance of lesion location and voxel-based lesion-symptom mapping

Relevance of lesion location: Lesion location is one important factor potentially contributing to the clinico-radiological paradox in MS or to the moderate WM lesion-symptom associations in cerebral SVD. Indeed, it is thought that while strategic lesions may directly lead to disability, other lesions in non-eloquent areas may remain clinically silent [5, 7]. For this reason, in the early 2000s, researchers started to implement different lesion mapping approaches, where not only the total lesion burden but also the anatomical location of lesions was respected [47].

Historically, mapping of lesions has been done without the use of objective statistics. For example, many studies have simply reported which region of the brain is most commonly damaged in individuals who exhibit a specific neurological impairment [48]. However, there are two main problems with this approach. First, it is not an objective test: Any random collection of patient lesions will necessarily show some region of maximal overlap, regardless of whether this location has any influence on the observed symptoms. Second, brain damage is not random: Because of the pathophysiological mechanisms underlying the disease, certain regions of the brain are particularly vulnerable to injury. Therefore, specific brain regions might commonly be injured regardless of their influence on symptoms [48].

For this reason, statistical lesion analyses to determine significant associations between symptoms and the location of brain injury were introduced. These analyses are either conducted by examining damage to predefined anatomical regions of interest or on a voxelwise basis [48]. Region of interest studies can offer better statistical power because fewer tests are computed and therefore less correction for multiple comparisons is needed. On the other hand, voxelwise lesion-mapping potentially offers better spatial resolution compared to region-based approaches and can reveal critical brain regions associated with a given deficit without *a priori* assumptions [48].

Voxel-based lesion-symptom mapping: In voxel-based approaches (the so-called voxel-based lesion-symptom mapping), the entire brain is mapped as a volume of voxels, with an independent statistical test conducted in each voxel [47]. A scheme summarizing a classical voxel-based lesion-symptom mapping pipeline is depicted in Figure 3. In a first step, lesions of all patients within a cohort are typically segmented (i.e. outlined) yielding an individual binary lesion mask for each patient. In a second step, these lesion masks are then transformed and overlaid (i.e. registered) on the same standard template so that anatomically homologous regions are brought into alignment [49]. Thirdly, for every voxel of this standard brain, it is computed whether or not injury to that voxel predicts a clinical deficit of interest in the given patient cohort. More in details, for each voxel, patients are divided into two groups according to whether they do have or do not have a lesion affecting that voxel. Clinical rating scores are then compared for these two groups using a statistical test suiting the nature of the clinical scale [47]. Finally, results are corrected for multiple comparisons and voxels that cross a given significance threshold are then considered as being part of the functionally critical brain areas whose damage leads to the deficit [49].

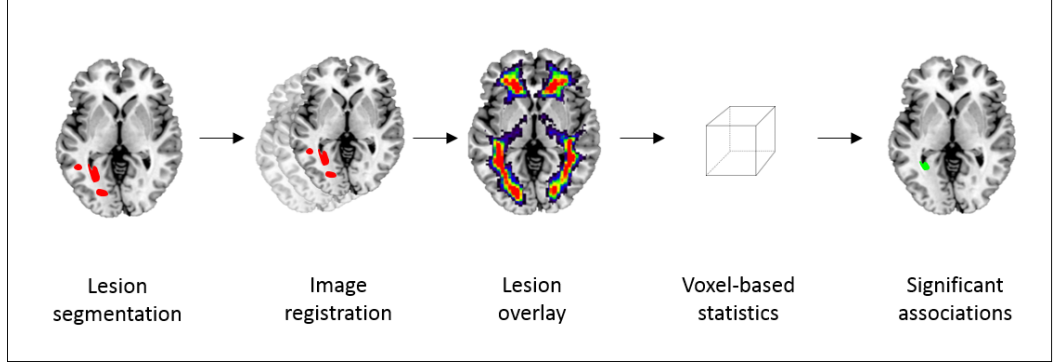


Fig. 3: Voxel-based lesion-symptom mapping. Scheme depicting the voxel-based lesion-symptom mapping pipeline. In a first step, lesions of all patients within cohort are segmented yielding an individual binary lesion mask for each patient. In a second step, these lesion masks are then aligned with or registered to a standard brain template. Then, for every voxel of this standard brain, it is computed whether or not injury to that voxel predicts a clinical deficit of interest in the given patient cohort. Finally, the statistical tests are corrected for multiple comparisons and a significance threshold is applied. In the final map, functionally relevant voxels are color-coded.

3.3 Voxel-based lesion-symptom mapping in multiple sclerosis

Several previous studies have investigated the association between strategic WM lesion location and standard clinical rating scales in MS using voxel-based approaches [8–14]. These studies and their most relevant findings are reported in Table 1. The study populations were either exclusively relapsing-remitting MS patients or a combination of different MS subtypes (including patients with clinically isolated syndrome). Most studies assessed disability using the two established clinical rating scales in MS (i.e. Expanded disability status scale, EDSS [50]; and MS functional composite, MSFC [51]; with/without separate correlations for functional subdomain scores), and a popular neuropsychological test battery (i.e. Brief repeatable battery of neuropsychological tests, BRB-N [52]; with separate correlations for all/some cognitive subdomain scores). Overall, studies agree that periventricular WM lesions non-specifically correlate with total EDSS and MSFC [8, 11, 12, 14]. These results can be explained by the fact that these clinical rating scales do not address a particular central nervous system function but comprise a multitude of clinical tests [11]. Moreover, in two studies, impaired memory could consistently be linked to lesions in the internal capsule [9, 10], which might highlight the relevance of thalamo-frontal connections for human memory. For the remaining findings, studies vary in the spatial resolution of the reported results and identified associations were often only reported by a single study.

Study	Domain	N	Lesions	Results
Charil et al., (2003), [8]	RRMS	452	WM (T1w, T2w)	<ul style="list-style-type: none"> - Integrative score: periventricular WM, internal capsule - Motor functions: internal capsule - Cognition: grey-white junction of associative, limbic, and prefrontal cortex - Vegetative functions: frontal lobe, parietal lobe, temporal lobe, dorsal midbrain, pons, cerebellum, insula - Coordination functions: periventricular WM - Brainstem functions: internal capsule - Sensory and visual functions: None
Sepulcre et al. (2008), [9]	MS	46	WM (T1w)	<ul style="list-style-type: none"> - Memory: temporal lobe (lateral to hippocampus and anterior temporal stem), thalamus, internal capsule, cingulum, fronto-occipital fasciculus
Sepulcre et al. (2009), [10]	MS	46	WM (T1w)	<ul style="list-style-type: none"> - Memory: cingulum, superior/inferior occipito-frontal fasciculus, longitudinal fasciculus, internal capsule, tegmentum, cerebellar peduncles
Vellinga et al.(2009), [11]	MS	325	WM (T2w)	<ul style="list-style-type: none"> - Integrative scores: Periventricular WM - Motor functions: Periventricular WM - Executive functioning: Periventricular WM
Kincses et al. (2011), [12]	MS, CIS	121	WM (T2w)	<ul style="list-style-type: none"> - Integrative score: periventricular WM - Motor, sensory, and coordination functions: None - Executive functioning: parietal lobe, superior longitudinal fasciculus (no results were found for the remaining cognitive scores tested)
Rossi et al. (2012), [13]	RRMS	142	WM (T2w)	<ul style="list-style-type: none"> - Processing speed: forceps major/minor, splenium of corpus callosum, inferior fronto-occipital fasciculus (no results were found for the remaining cognitive scores tested)
Ge et al. (2014), [14]	MS, CIS	250	WM (T2w)	<ul style="list-style-type: none"> - Integrative score: periventricular WM, forceps minor/major - Processing speed: periventricular WM, corpus callosum

Tab. 1: Summary on previous voxel-based lesion-symptom mapping studies in MS. Table reporting the identified associations between MS lesion location and disease symptoms as assessed using established clinical rating scales in MS. Only results corrected for multiple

comparisons are reported. N: Number of enrolled patients; RRMS: Relapsing-remitting MS; CIS: clinically isolated syndrome; T1w: T1-weighted imaging sequence; T2w: T2-weighted imaging sequence.

3.4 Voxel-based lesion-symptom mapping in cerebral small vessel disease

More recently, the first voxel-based and region of interest-based lesion-symptom mapping studies have also emerged in SVD (Table 2) [15, 17, 19]. The study populations varied from healthy individuals, patients with manifest arterial disease, patients with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL, a rare monogenetic form of SVD), to memory clinic patients. These studies thus cover the full spectrum of SVD-related cognitive impairment, which ranges from subtle cognitive disturbances to manifest dementia, from pre-clinical MRI manifestations of SVD to extensive WM damage, from pure vascular aetiology (such as CADASIL) to mixed pathologies. Most studies focused on processing speed and executive functions, which are typically the most severely affected in patients with SVD, while some studies also included memory as outcome measure. These studies consistently found WM lesions or lacunar volume in the anterior thalamic radiation and the forceps minor to impact on poor processing speed [15, 17, 19]. The anterior thalamic radiation connects thalamic nuclei with frontal and cingulate cortices and cognitive impairment most likely results from a disconnection syndrome, which highlights the crucial role of frontal-subcortical projections in SVD-related impairment in processing speed. Then, studies reported a relationship between WM lesions and lacunar volume in the anterior thalamic radiation, forceps minor and superior longitudinal fasciculus and poor executive functioning [16, 19]. The superior longitudinal fasciculus is a major pathway that connects frontal, parietal, and temporal association cortices and is known to be relevant in executive functions [53]. Associations between memory and WM lesions were found in forceps minor, forceps major, temporo-occipital WM, right parietal WM, and the left internal capsule, though there is insufficient evidence to draw definite conclusions to date [15, 18, 19].

Study	Domain	N	Lesions	Results
Duering et al. (2011), [15]	CADASIL	215	WM (T2w), Lacunes	<ul style="list-style-type: none"> - Processing speed: anterior thalamic radiation, forceps major, forceps minor, corticospinal tract - Memory: forceps major
Smith et al. (2015), [18]	Controls, MCI, AD	145	WM (T2w)	<ul style="list-style-type: none"> - Executive functioning: inferior frontal and temporal-occipital WM, parietal WM, anterior internal capsule - Memory: inferior temporo-occipital WM, temporo-occipital WM, right parietal WM, anterior internal capsule
Biesbroek et al. (2013), [16]	Manifest arterial disease	516	WM (T2w), Lacunes	<ul style="list-style-type: none"> - Processing speed: none - Executive functioning: anterior thalamic radiation, superior longitudinal fasciculus - Memory: none
Duering et al. (2014), [17]	Controls	584	WM (T2w)	<ul style="list-style-type: none"> - Processing speed: left anterior thalamic radiation, forceps minor
Biesbroek et al. (2016), [19]	Memory clinic patients with SVD	167	WM (T2w)	<ul style="list-style-type: none"> - Processing speed: anterior thalamic radiation, forceps minor - Executive functioning: anterior thalamic radiation, forceps minor - Memory: forceps minor

Tab. 2: Summary on previous voxel-based lesion-symptom mapping studies in SVD. Table reporting the identified associations between WM lesion or lacune location and cognition. N: Number of enrolled patients; CADASIL: cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy; MCI: Mild cognitive impairment; AD: Alzheimer’s disease; T2w: T2-weighted imaging sequence.

4 Two large cohort voxel-based lesion-symptom mapping studies

Voxel-based lesion-symptom mapping was performed in two large cohorts: a dataset of 2348 relapsing-remitting MS patients (from the multicenter FREEDOMS and FREEDOMS II phase 3 clinical trials; first publication presented in this chapter), and a population-based cohort of 1017 elderly subjects (from the INtervention project on cerebroVascular disease and Dementia in the district of Ebersberg, INVADE; second publication presented in this chapter). Prior to our analyses, lesion segmentation of T2-weighted WM lesions (in the dataset of MS patients) and WMHs (in the population-based cohort of elderly subjects) was done at MIAC AG (Medical Image Analysis Center AG). In a first step, for both studies, an automatized registration pipeline for alignment of the provided binary lesion masks to standard stereotaxic space was elaborated and implemented. In a second step, the normalized lesion masks were overlaid, and a voxel-based statistical approach to relate lesion location to different clinical measures was defined and encoded.

Publication: *Clinical correlations of brain lesion location in multiple sclerosis: voxel-based analysis of a large clinical trial dataset.* Anna Altermatt, Laura Gaetano, Stefano Magon, Dieter A. Häring, Davorka Tomic, Jens Wuerfel, Ernst-Wilhelm Radue, Ludwig Kappos, Till Sprenger. This publication has been submitted to the journal Brain Topography and is currently under review.

Publication: *Association of cerebral T2-weighted lesion load and location, vascular risk factors, and cognition in a large population-based cohort.* Anna Altermatt, Laura Gaetano, Stefano Magon, Lorena Bauer, Regina Feurer, Hans Gnahn, Julia Hartmann, Christian L. Seifert, Holger Poppert, Jens Wuerfel, Ernst-Wilhelm Radue, Ludwig Kappos, Till Sprenger. This publication has been submitted to the journal NeuroImage and is currently under review.

Clinical correlations of brain lesion location in multiple sclerosis: voxel-based analysis of a large clinical trial dataset

Anna Altermatt ^{1,2}, Laura Gaetano ^{2,3}, Stefano Magon ^{2,3}, Dieter A. Häring ⁵, Davorka Tomic ⁵, Jens Wuerfel ^{1,2}, Ernst-Wilhelm Radue ⁶, Ludwig Kappos ³, Till Sprenger ^{2,3,4}

¹ Department of Biomedical Engineering, University of Basel, 4123 Allschwil, Switzerland

² Medical Image Analysis Center (MIAC) AG, Mittlere Strasse 83, 4056 Basel, Switzerland

³ Department of Neurology, University Hospital of Basel, Petersgraben 4, 4031 Basel, Switzerland

⁴ Department of Neurology, DKD HELIOS Klinik Wiesbaden, Aukammallee 33, 65191 Wiesbaden, Germany

⁵ Novartis Pharma AG, Postfach, 4002 Basel, Switzerland

⁶ Biomedical Research and Education GmbH, Mittlere Strasse 83, 4056 Basel, Switzerland

Corresponding author: Anna Altermatt; Email: anna.altermatt@unibas.ch; Phone: +41 61 265 43 18

Abstract

There is a limited correlation between white matter (WM) lesion load as determined by magnetic resonance imaging and disability in Multiple Sclerosis (MS). The reasons for this so-called clinico-radiological paradox are diverse and may, at least partly, relate to the fact that not just the overall lesion burden, but also the exact anatomical location of lesions predict the severity and type of disability. We aimed at studying the relationship between lesion distribution and disability using a voxel-based lesion probability mapping approach in a very large dataset of MS patients. T2-weighted lesion masks of 2348 relapsing-remitting MS patients were spatially normalized to standard stereotaxic space by non-linear registration. Relations between supratentorial WM lesion locations and disability measures were assessed using a non-parametric ANCOVA (Expanded Disability Status Scale, EDSS; Multiple Sclerosis Functional Composite, MSFC, and subscores; Modified Fatigue Impact Scale, MFIS) or multinomial ordinal logistic regression (EDSS functional subscores). Data from 1907 (81%) patients were included in the analysis because of successful registration. The lesion mapping showed similar areas to be associated with the different disability scales: periventricular regions in temporal, frontal, and limbic lobes were predictive, mainly affecting the posterior thalamic radiation, the anterior, posterior, and superior parts of the corona radiata. In summary, significant associations between lesion location and clinical scores were found in periventricular areas. Such lesion clusters appear to be associated with impairment of different physical and cognitive abilities, probably because they affect commissural and long projection fibers, which are relevant WM pathways supporting many different brain functions.

Keywords: Multiple sclerosis; magnetic resonance imaging; motor impairment; cognitive dysfunction; lesion probability map; white matter

Introduction

Although conventional magnetic resonance imaging (MRI) is sensitive for depicting white matter (WM) lesions in Multiple Sclerosis (MS), the association between the cumulative WM lesion load and clinical disability is at best moderate (Barkhof, 2002). The reasons of this so-called clinico-radiological paradox are probably diverse, ranging from methodological to physiological aspects (Mollison et al. 2017). Lesion location is one factor potentially contributing to this paradox: while strategic lesions may directly lead to disability, other (even larger) lesions in non-eloquent areas may remain clinically silent. Hence, to better predict clinical scores, one would need to take lesion location into account (Kincses et al. 2011). In this regard, several previous studies have investigated the relation between WM lesion distribution and neurological deficits in patients with different subtypes of MS using voxel-based approaches (Charil et al. 2007; Sepulcre et al. 2008; Sepulcre et al. 2009; Vellinga et al. 2009; Kincses et al. 2011; Rossi et al. 2012). In a larger study by Charil et al. (2007), significant correlations to motor scores were found to be located mainly in the internal capsule. Moreover, it has been shown that lesions in WM tracts connecting associative areas are correlated with cognitive impairment. Another group (Kincses et al. 2011) was able to link compromised sensory functions to lesions in the left thalamus, while cerebellar lesions were shown to account for deficits in coordination. In the same study, amygdalar, frontal, temporal, and parietal lesion locations were found to correlate with cognitive performance. Rossi et al. (2012) showed that patients with MS and cognitive impairment more frequently had lesions in commissural fiber tracts. They hypothesized that a functional disconnection between key grey matter structures may (partly) account for cognitive impairment in MS. Sepulcre et al. (2008) reported lesion locations correlating with declarative verbal memory storage in the temporal lobe (particularly in the anterior temporal stem), the left thalamus, and the left internal capsule. The same volumes were shown to be relevant for retrieval, but to them were added the cingulum and the fronto-occipital fasciculus. In a further study (Sepulcre et al. 2009), the same group described significant associations of a verbal memory task with lesions in the cingulum, in parieto-frontal pathways, and within thalamo-cortical projections as well as in the right cerebellar WM. Finally, Vellinga et al. (2009) identified bilateral periventricular clusters to correlate with different disability measures.

Most previous lesion mapping studies in MS included relatively small numbers of patients with only few studies that enrolled more than hundred patients (Charil et al. 2007; Vellinga et al. 2009; Kincses et al. 2011; Rossi et al. 2012). Most studies considered few clinical scores or only composite scores of disability (Vellinga et al. 2009). Moreover, in several previous studies individual lesions (in subject space) were transformed to standard stereotaxic space using linear transformations only (Charil et al. 2003; Vellinga et al. 2009), potentially introducing bias regarding the exact anatomical location in standard space (Klein et al. 2009). Finally, in a few studies (Charil et al. 2003), the binary nature of the lesion maps was not respected (Ge et al. 2014), as conventional parametric statistical models were applied.

Therefore, the objective of this study is to investigate the relation between supratentorial WM lesion location and 13 different disability scales in a large clinical trial dataset. Investigated clinical scores comprise multiple domains of impairment seen in MS and include EDSS (Expanded Disability Status Scale and functional subscores; Kurtzke, 1983) with scores of functional subsystems, MSFC (Multiple Sclerosis Functional Composite; Fischer et al. 1999) with subscores, and MFIS (Modified Fatigue Impact Scale; Fisk et al. 1994). Alignment to standard stereotaxic space is done using non-linear registration. Voxel-by-voxel associations between lesion occurrence and

clinical disability are assessed using a non-parametric ANCOVA or ordinal logistic regression. Furthermore, principally affected WM tracts overlapping with statistically significant locations are identified.

Materials and Methods

2.1 Patients

All available data from patients participating in the multicenter FREEDOMS (N=1272) and FREEDOMS II (n=1083) phase 3 clinical trials (details in Kappos et al. 2010; Calabresi et al. 2014) were analyzed. A total of 2355 relapsing-remitting MS (RRMS) patients, diagnosed according to the revised McDonald criteria (Polman et al. 2005), were analyzed from these two studies. In the trials, MRI scans were conducted at baseline, month 6, month 12 and month 24, but only baseline data were considered in this lesion-mapping study.

2.2 Disability scores

EDSS (Kurtzke 1983) with scores of functional subsystems (pyramidal, sensory, cognition, vegetative, visual), MSFC (Fischer et al. 1999) including sub-scores (Timed 25-Foot Walk, T25FW; 9-Hole Peg Test, NHPT; Paced Auditory Serial Addition Test, PASAT) and MFIS (Fisk et al. 1994) data were available.

2.3 MR imaging protocol

The MRI protocol of the two studies included T1-weighted (T1w) images and a dual echo proton density-weighted (PDw)/T2-weighted (T2w) sequence (in plane resolution = 1x1 mm², slice thickness = 3 mm) acquired at 1.0 (3.6%), 1.5 (87.8%), or 3.0 (8.6%) Tesla (T) scanners depending on the local MRI center. In 5/2355 (<1%) patients, PDw/T2w images were missing or of low quality and subsequent lesion segmentation could not be carried out. In 1/2355 (<1%) patient, the T1w images were missing, leading to an exclusion of this subject as subsequent image registration to standard stereotaxic space was not possible.

2.4 Lesion segmentation

Lesion segmentation was done by trained and experienced raters at MIAC AG (Medical Image Analysis Center AG, Basel, Switzerland) in three distinct steps according to the in-house standard operating procedures. In a first step, lesions are assessed and marked according to the trial-specific evaluation matrix and using the software package AMIRA 3.1.1 (Mercury Computer System Inc.). The results of the marking process are checked by the rater responsible for the following segmentation step (first consensus reading). Lesions are then automatically segmented on the PDw images using the same software and manually corrected by the second rater if needed. All raters undergo a training period and a standardized reliability testing prior to any study involvement. Reliability is tested in all raters at yearly intervals. This ensures a consistent quality of lesion marking and segmentation. After lesion marking and segmentation, results are reviewed by a radiologist (second consensus reading). In case of discrepancies in any of the consensus reading, the examination is returned to the previous rater to achieve an agreement (Magon et al. 2014).

2.5 Image registration

Binary lesion masks were first registered to T1w images applying the transformation parameters resulting from linearly registering the PDw image to the T1w image using FLIRT (FMRIB's linear registration tool; default settings, 6 degrees of freedom; Jenkinson and Smith, 2001; Jenkinson et al. 2002). These lesion masks were used to perform lesion filling on the T1w images (Sdika and Pelletier, 2009; Magon et al. 2014). Then, a two-stage linear (default

settings, 6 degrees of freedom) and subsequently non-linear registration was carried out to align the T1w images to the MNI152 standard brain template (2x2x2 mm³) (Mazziotta et al. 2001) using FNIRT (settings as specified in the provided configuration file, Jacobian range between 0.1 – 10; Andersson et al. 2007). The obtained transformation matrices were applied to the lesion masks previously registered to the T1w images. The quality assessment of the registration was done visually in two steps by two trained and experienced raters. In a first step, registrations were either approved or rejected by a first rater based on the quality of the aligned images. In a second step, all images were again rated independently by a second person. In case of discrepancies, a common agreement was found between the two raters (consensus reading). Using this procedure, 442/2355 (19%) patients were excluded from subsequent statistical analyses due to distortions in the registered images. Finally, the transformed binary lesion maps of all accepted subjects were overlaid in MNI152 space and a lesion distribution map was computed.

2.6 Statistical analysis

For each voxel, patients were divided into two groups depending on whether they did or did not have a lesion in this voxel (Bates et al. 2003). Disability scores were then compared for those two groups. A non-parametric ANCOVA was employed for the main disability scores (i.e. EDSS, MSFC, T25FW, PASAT, NHPT, MFIS), while a multinomial ordinal logistic regression model was employed for the EDSS functional subsystem scores (distinguishing maximally up to 6 scoring levels). In the latter model, which applies parametric statistics, a minimal lesion occurrence threshold of 30 lesions per voxels was set. This minimal lesion occurrence (i.e. event rate) for the logistic regression model was estimated to allow for the detection of a difference of 25% in the number of subjects with an increased disability score in the subjects with a lesion in a voxel compared to the subject without a lesion in the same voxel. Based on a power of 0.8 to detect a significant difference of this effect size and a 5% significance level, a minimum of 30 patients with a lesion in a voxel were required. The models were corrected for the effects of age, gender, disease duration, imaging center, and field strength. All statistical tests were adjusted for multiplicity applying a Benjamini & Hochberg procedure ($\alpha=0.05$) (Benjamini and Hochberg 1995). The whole statistical analysis was implemented in R (<http://r-project.org/>). Finally, the percentage of voxels showing significant correlations between disability measures and lesion location was computed (i.e. percentage of significant voxels, $PSV = 100 \times \text{number of voxels significant} / \text{total number of voxels affected by lesions}$) and the WM fiber tracts affected were identified using the JHU white-matter labels atlas (Mori and Crain, 2005) available in FSL.

Results

3.1 Demographic data, MS disease characteristics, and clinical test performance

Following successful T2w lesion segmentation and registration to MNI space, a total of 1907/2355 (81%) patients were included in the voxel-wise analysis. Clinical data are represented in Table 1.

Table 1: Demographic data, disease features, and clinical test performance

	Median (range) or N	N (%)
<i>Demographics</i>		
Age (years)	39 (17-57)	1907 (100%)
Women	1362	1907 (100%)
<i>MRI-related features</i>		
Imaging centers (FREEDOMS/FREEDOMS II)	138/117	1907 (100%)
Field strength (1T/1.5T/3T)	84/2068/203	1907 (100%)
<i>Disease features</i>		
Disease duration (years)	7.3 (0.2-49.9)	1907 (100%)
Lesion load (T2w lesion volume) (mm ³)	2936 (0-69203)	1907 (100%)
<i>Clinical test performance</i>		
EDSS	2 (0-6)	1907 (100%)
Pyramidal (pyramidal functions)	1 (0-5)	1907 (100%)
Sensory (sensory functions)	1 (0-5)	1907 (100%)
Cognition (cerebral functions)	0 (0-3)	1907 (100%)
Vegetative (bowel & bladder functions)	0 (0-4)	1907 (100%)
Visual (visual functions)	0 (0-4)	1907 (100%)
MSFC (z-score)	0.1388 (-6.4133-2.0639)	1886 (98.9%)
T25FW	5.1 (2.1-91.5)	1899 (99.6%)
NHPT	20.8 (9.2-330.5)	1891 (99.2%)
PASAT	52 (0-60)	1894 (99.3%)
MFIS	23 (0-63)	790 (41.4%)

N=number of patients. EDSS: Expanded Disability Status Scale and functional subscores; MSFC: Multiple Sclerosis Functional Composite; T25FW: Timed 25-Foot Walk; NHPT: 9-Hole Peg Test; PASAT: Paced Auditory Serial Addition Test; MFIS: Modified Fatigue Impact Scale.

3.2 Spatial distribution of MS lesions

The lesion masks of 1907 patients were summed in MNI space (Fig. 1) showing a scattered distribution of lesions in the WM, with the highest lesion occurrence in the periventricular areas of both hemispheres. In those regions, individual voxel reached a maximal lesion overlap of 420 (patients), meaning that up to 22% of patients were characterized by lesions in a given voxel.

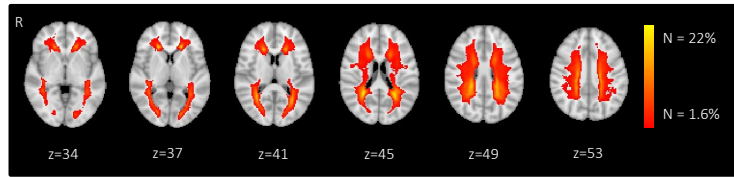


Figure 1: Lesion distribution. Voxelwise lesion distribution across the patient group in stereotaxic standard space. Only voxels where at least 30/1907 (1.6%) patients had a lesion are color-coded.

3.3 Association of lesion location with clinical scale

An overview of the correlational lesion mapping results, by clinical scale, is shown in Fig. 2. Voxel counts in both hemispheres and PSV for each score are depicted in Table 2. Table 3 lists the WM fiber tracts coinciding with lesional clusters that reached statistical significance. The highest PSV was reached for MSFC (18.4%), NHPT (18.7%) and PASAT (11.8%); the according voxels were located mainly in the WM of the temporal, frontal, and limbic lobe, interrupting wide parts of the corpus callosum, bilaterally the anterior, superior and posterior corona radiata, and the posterior thalamic radiation in both hemispheres. Fewer correlating lesional clusters were identified for the overall EDSS (PSV = 5.8%), pyramidal EDSS subscore (PSV = 6.9%), and the T25FW (PSV = 4.8%). The spatial distribution found for these scores was very similar, affecting a subarea of the regions found for the scores with highest PSV. Most affected WM tracts for the latter scores were bilaterally distributed in the anterior, superior and posterior corona radiata and the posterior thalamic radiation. The most restricted and well-defined patterns of spatial distribution were found for the visual EDSS subscore (PSV = 1.2%) and MFIS (PSV = 0.5%). For the visual EDSS subscore, the right posterior corona radiata and the left thalamic radiation (including optic radiation) were the most affected WM tracts. For the MFIS, significant voxels were distributed posteriorly and coinciding mainly with the posterior corona radiata of both hemispheres. No correlating brain regions were found for sensory, cognitive, and vegetative EDSS subscores.

Table 2: Number of voxels reaching statistical significance and PSV

	Number of significant voxels			PSV (%)
	Left	Right	Total	Total
EDSS	3034	3282	6316	5.8
Pyramidal	3612	3949	7561	6.9
Visual	683	612	1295	1.2
MSFC	10327	9777	20104	18.4
T25FW	2496	2767	5263	4.8
NHPT	10224	10202	20426	18.7
PASAT	6613	6267	12880	11.8
MFIS	229	324	553	0.5

EDSS: Expanded Disability Status Scale and functional subscores; MSFC: Multiple Sclerosis Functional Composite; T25FW: Timed 25-Foot Walk; NHPT: 9-Hole Peg Test; PASAT: Paced Auditory Serial Addition Test; MFIS: Modified Fatigue Impact Scale. "Left" and "right" refers to the left and right cerebral hemispheres. PSV refers to the percentage of significant voxels and is calculated as $PSV = 100 \times \text{number of voxels reaching significance} / \text{number of voxels affected by lesions}$. The total number of voxels affected by lesions corresponds to all voxels where at least one lesion occurred ($N=109183$).

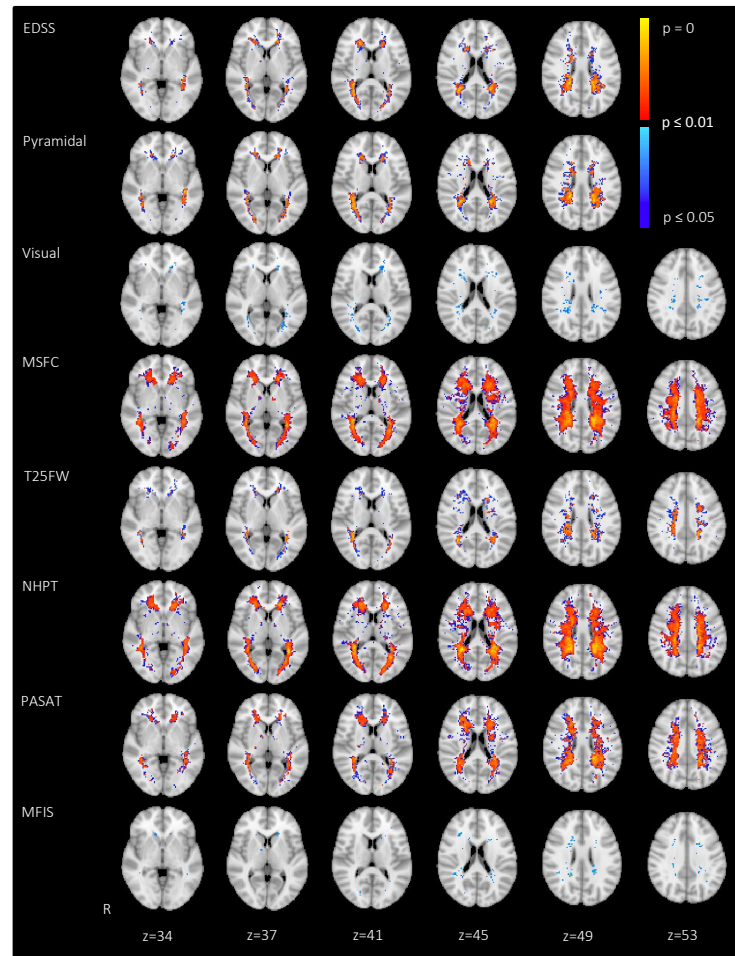


Figure 2: Significant correlations between lesion location and disability scores. The maps depict voxels with statistically significant ($p < 0.05$ in blue; $p < 0.01$ in red) association to disability scores after applying the Mann-Whitney U test (in case of EDSS, MSFC, T25FW, NHPT, PASAT, MFIS) or ordinal regression (for pyramidal and visual EDSS subscores) in standard stereotaxic space. EDSS: Expanded Disability Status Scale and functional subscores; MSFC: Multiple Sclerosis Functional Composite; T25FW: Timed 25-Foot Walk; NHPT: 9-Hole Peg Test; PASAT: Paced Auditory Serial Addition Test; MFIS: Modified Fatigue Impact Scale.

Table 3: Localization of lesional associations with clinical scores

WM tracts	Percentage of WM tract affected by lesional voxels that were associated with the respective score								Percentage of significant voxels on WM tract							
	EDSS	Pyr.	Visual	MSFC	T25FW	NHPT	PASAT	MFIS	EDSS	Pyr.	Visual	MSFC	T25FW	NHPT	PASAT	MFIS
Genu of corpus callosum	15.7	17.4	4.5	35.2	8.9	33.1	26.2	1.1	2.8	2.6	3.9	2.0	1.9	1.8	2.3	2.2
Body of corpus callosum	11.4	16.5	0.6	30.2	9.0	30.9	19.7	1.5	3.1	3.8	0.8	2.6	3.0	2.6	2.6	4.7
Splenium of corpus callosum	14.7	16.2	1.9	24.8	9.7	27.5	15.9	2.1	3.6	3.3	2.2	1.9	2.9	2.1	1.9	6.0
Anterior corona radiata R	35.9	32.9	6.9	77.2	25.4	79.1	64.0	2.8	4.9	3.7	4.6	3.3	4.1	3.3	4.3	4.3
Anterior corona radiata L	29.2	32.0	7.7	81.7	27.2	80.5	63.4	2.4	4.0	3.7	5.2	3.5	4.5	3.4	4.3	3.8
Superior corona radiata R	31.2	46.8	3.9	89.5	39.5	87.1	70.8	3.8	4.5	5.7	2.8	4.1	6.9	3.9	5.1	6.3
Superior corona radiata L	27.9	35.6	4.8	87.7	33.3	80.2	75.8	2.9	4.1	4.4	3.4	4.0	5.9	3.6	5.4	4.9
Posterior corona radiata R	81.0	86.7	23.2	98.7	77.7	98.5	92.7	10.4	5.8	5.2	8.1	2.2	6.7	2.2	3.3	8.5
Posterior corona radiata L	81.2	84.5	12.8	96.9	65.9	97.3	92.6	11.0	5.7	5.0	4.4	2.1	5.6	2.1	3.2	8.9
Posterior thalamic radiation (optic rad.) R	68.4	80.7	16.8	94.7	61.8	93.2	74.3	1.8	5.3	5.2	6.3	2.3	5.7	2.2	2.8	1.6
Posterior thalamic radiation (optic rad.) L	59.4	68.8	24.9	95.6	51.0	95.4	79.9	1.3	4.5	4.4	9.2	2.3	4.6	2.2	3.0	1.1
Superior longitudinal fasciculus L	10.7	16.2	5.5	70.4	14.0	69.0	44.7	0.1	1.4	1.7	3.5	2.9	2.2	2.8	2.8	0.2

WM tract computations were done using JHU WM atlas labels (Mori et al. 2005) available in FSL. Only WM tracts where at least 3% of the significant voxels were located for one disability score, are shown. EDSS: Expanded Disability Status Scale and functional subscores (pyramidal and visual); MSFC: Multiple Sclerosis Functional Composite; T25FW: Timed 25-Foot Walk; NHPT: 9-Hole Peg Test; PASAT: Paced Auditory Serial Addition Test; MFIS: Modified Fatigue Impact Scale.

Discussion

We studied the correlation between supratentorial lesion location and clinical scores in a large cohort of RRMS patients. Voxel-by-voxel associations between lesion occurrence and clinical impairment were assessed using either a non-parametric ANCOVA or ordinal logistic regression and WM tracts overlapping with statistically significant locations were identified.

Significant associations between lesion location and clinical scores were mainly found in periventricular areas, reflecting the typical lesion distribution seen in MS. Periventricular lesions principally affect commissural and long-distance projection fibers and damage to even a small number of such large WM tracts might have a dramatic effect on different brain networks causing a variety of disease symptoms (Sepulcre et al. 2009). The size of correlating clusters within these areas depends on the investigated scale: MSFC, NHPT and PASAT showed largest clusters of correlating voxels. Fewer correlating lesional clusters were identified for the overall EDSS, pyramidal EDSS, and the T25FW. Results found for the MSFC seem to resemble the widespread patterns shown by the NHPT and PASAT. Similarly, results found for the total EDSS, seem to be dominated by the pyramidal EDSS subscore. When comparing results of total MSFC and EDSS, we could speculate that the tasks involved in the MSFC subscores seem to recruit a more complex network of brain regions suggesting that the latter score captures the MS pathology more holistically. A potential explanation to this finding could lie in the fact that the MSFC, in addition to motor abilities, emphasizes more on the cognitive aspects of MS (Fischer et al. 1999) when compared to the EDSS (Kurtzke 1983). The most well-defined clusters were found for the visual EDSS subscore, and the MFIS. Lesion mapping using visual EDSS subscores revealed focal clusters in the posterior corona radiata and the posterior thalamic radiation, coinciding with the optic radiation. MFIS correlations were mainly restricted to the posterior corona radiata of both hemispheres.

Our results indicate a limited specificity in terms of distinctive regions being related to different scores. This is in accordance with the findings of Vellinga et al. (2009), who observed rather diffuse patterns of correlation between lesion location and EDSS, MSFC and its subscores in periventricular areas. Some other previous studies (Sepulcre et al. 2008; Sepulcre et al. 2009; Kincses et al. 2011; Rossi et al. 2012) showed a more focal relationship between lesion location and disability measures, though none of them reported spatial correlations with pyramidal and visual EDSS subscores. For example, Kincses et al. (2011) were able to link lesions in the left thalamus and cerebellum, respectively, to sensory and coordination deficits. However, most results were only significant if uncorrected statistics were employed. Rossi et al. (2012) stated that areas relevant for cognition are located in commissural fiber tracts. Sepulcre et al. (2008) reported lesion locations correlating with declarative verbal memory storage in the temporal lobe (particularly in the anterior temporal stem), the left thalamus, the left internal capsule, the cingulum, and the fronto-occipital fasciculus. In a further study (Sepulcre et al. 2009), the same group described significant associations of a verbal memory task with lesions in the cingulum, in parieto-frontal pathways, and within thalamo-cortical projections, and in the right cerebellar WM. However, this group studied lesion-symptom associations in a comparatively small dataset (Sepulcre et al. 2008; Sepulcre et al. 2009). Therefore, previous results might be driven by trends specific to the dataset rather than being representative for MS in general.

The relative lack of specific WM areas related to some of the disability measures we observed could be attributed to several reasons. First of all, the scales used in this analysis assess relatively complex brain functions that are probably not related to one specific region or network in the brain, but rather reflect multifocal neural systems (Bressler and Tognoli 2006). For example, functional MRI experiments conducted during PASAT performance in

healthy subjects have shown that a complex pattern of parietal, frontal, limbic, occipital, and temporal cortical activation occurs during this demanding verbal working memory/processing speed task (Tombaugh 2006). Complex activation patterns have also been shown related to even relatively simple motor tasks. As an example, the T25FW has been shown to be associated with networks encoding attention and executive function (Benedict et al. 2011). This suggests that disability as measured by the applied scores results from pathway disruptions at any possible location within complex brain networks involved in the respective task. Another important aspect that could influence the relative lack of specificity observed in our results could relate to compensatory strategies to cope with functional/behavioral deficits and neuroplastic mechanisms. Indeed, mechanisms of adaptive reorganization or plasticity limits the effects of brain damage on daily life and, as a consequence, may impact the scores of disability scales (Tomassini et al. 2012).

We acknowledge that the relative lack of specific areas associated with different disability measures could also relate to technical factors: For a given dataset of this size, very small group differences may be detectable in voxels with high lesion frequency (Sullivan and Feinn 2012). Imposing a significance threshold may help to identify areas that are correlated with a clinical score with high certainty, but it may also limit the ability to observe signals in areas where lesions occur less frequently. Significant correlations are therefore more likely found in those voxels where lesions are commonly present in MS patients. In addition, the applied analysis method considers each voxel independently and cannot borrow information from neighboring voxels in the same brain region, which may have limited our ability to detect association between lesions in regions that are infrequently occupied by lesions (e.g. in the infratentorial brain) and clinical scales (e.g. walking ability). Moreover, the type of lesions considered for this study could influence the results: T2w lesions reflect a variety of pathological processes (i.e. acute inflammation, edema, demyelination, remyelination, axonal loss etc.) that may impact MS-related disability in different ways. The cross-sectional assessment of T2w lesions represents the history of a patient's MS, and a single MRI scan does not allow differentiation of the severity of tissue damage (Rovira et al. 2013). Finally, lesions in other CNS regions known to be relevant for the development of disability were not included in the analyses. Although MS is commonly termed chronic inflammatory-demyelinating disease of the WM, several recent neuropathological studies disclosed an extensive involvement of the cortex and deep GM, where the extend is also correlated to physical disability and cognitive dysfunction (Calabrese et al. 2013).

In conclusion, we found significant associations between lesion location and different clinical scores in periventricular areas. This reflects the characteristic lesion distribution seen in MS. It is likely that periventricular lesions affect commissural and long projection fibers, which are involved in a multitude of brain functions. Results found for the MSFC seem resemble patterns identified for NHPT, and to a lesser extent, for the PASAT. Similarly, results found for the total EDSS, seem to be dominated by the pyramidal EDSS subscore. The relative numbers of voxels showing significant associations for the main clinical scores suggest that the MSFC recruits a more complex network of brain regions when compared to the EDSS possibly due to the stronger weighting of cognition in the MSFC.

References

- Andersson J, Jenkinson M, Smith S (2007) Non-linear registration aka s spatial normalisation FMRI Technical Report TR07JA2
- Barkhof F (2002) The clinico-radiological paradox in multiple sclerosis revisited. *Curr Opin Neurol* 15:239–45
- Bates E, Wilson SM, Saygin AP, et al (2003) Voxel-based lesion–symptom mapping. *Nat Neurosci* 6:448–50. doi: 10.1038/nn1050
- Benedict RHB, Holtzer R, Motl RW, et al (2011) Upper and Lower Extremity Motor Function and Cognitive Impairment in Multiple Sclerosis. *J Int Neuropsychol Soc* 17:643–653. doi: 10.1017/S1355617711000403
- Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B* 57:289–300
- Bressler SL, Tognoli E (2006) Operational principles of neurocognitive networks. *Int J Psychophysiol* 60:139–148. doi: 10.1016/j.ijpsycho.2005.12.008
- Calabrese M, Favaretto A, Martini V, Gallo P (2013) Grey matter lesions in MS: from histology to clinical implications. *Prion* 7:20–7. doi: 10.4161/pri.22580
- Calabresi PA, Radue E-W, Goodin D, et al (2014) Safety and efficacy of fingolimod in patients with relapsing-remitting multiple sclerosis (FREEDOMS II): a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Neurol* 13:545–556. doi: 10.1016/S1474-4422(14)70049-3
- Charil A, Zijdenbos AP, Taylor J, et al (2003) Statistical mapping analysis of lesion location and neurological disability in multiple sclerosis: application to 452 patient data sets. *Neuroimage* 19:532–544. doi: 10.1016/S1053-8119(03)00117-4.
- Fischer JS, Rudick RA, Cutter GR, Reingold SC (1999) The Multiple Sclerosis Functional Composite measure (MSFC): an integrated approach to MS clinical outcome assessment. *Mult Scler J* 5:244–250. doi: 10.1177/135245859900500409
- Fisk JD, Ritvo PG, Ross L, et al (1994) Measuring the functional impact of fatigue: initial validation of the fatigue impact scale. *Clin Infect Dis* 18:79–83
- Ge T, Müller-Lenke N, Bendfeldt K, et al (2014) Analysis of multiple sclerosis lesions via spatially varying coefficients. *Ann Appl Stat* 8:1095–1118. doi: 10.1214/14-AOAS718
- Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17:825–41. doi: 10.1006/nimg.2002.1132
- Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. *Med Image Anal* 5:143–56. doi: 10.1016/S1361-8415(01)00036-6
- Kappos L, Radue E-W, O'Connor P, et al (2010) A Placebo-Controlled Trial of Oral Fingolimod in Relapsing Multiple Sclerosis. *N Engl J Med* 362:387–401. doi: 10.1056/NEJMoa0909494
- Kincses ZT, Ropele S, Jenkinson M, et al (2011) Lesion probability mapping to explain clinical deficits and cognitive performance in multiple sclerosis. *Mult Scler* 17:681–9. doi: 10.1177/1352458510391342
- Klein A, Andersson J, Ardekani BA, et al (2009) Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46:786–802. doi: 10.1016/j.neuroimage.2008.12.037

- Kurtzke JF (1983) Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology* 33:1444–1453. doi: 10.1212/WNL.33.11.1444
- Magon S, Gaetano L, Chakravarty MM, et al (2014) White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: a longitudinal study. *BMC Neurosci* 15:106. doi: 10.1186/1471-2202-15-106
- Mazziotta J, Toga A, Evans A, et al (2001) A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos Trans R Soc Lond B Biol Sci* 356:1293–322. doi: 10.1098/rstb.2001.0915
- Mollison D, Sellar R, Bastin M, et al (2017) The clinico-radiological paradox of cognitive function and MRI burden of white matter lesions in people with multiple sclerosis: A systematic review and meta-analysis. *PLoS One* 12:e0177727. doi: 10.1371/journal.pone.0177727
- Mori S, Crain BJ (2005) *MRI Atlas of Human White Matter*. Elsevier, Amsterdam
- Polman CH, Reingold SC, Edan G, et al (2005) Diagnostic criteria for multiple sclerosis: 2005 revisions to the “McDonald Criteria.” *Ann Neurol* 58:840–846. doi: 10.1002/ana.20703
- Rossi F, Giorgio A, Battaglini M, et al (2012) Relevance of brain lesion location to cognition in relapsing multiple sclerosis. *PLoS One* 7:e44826. doi: 10.1371/journal.pone.0044826
- Rovira A, Auger C, Alonso J (2013) Magnetic resonance monitoring of lesion evolution in multiple sclerosis. *Ther Adv Neurol Disord* 6:298–310. doi: 10.1177/1756285613484079
- Sdika M, Pelletier D (2009) Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. *Hum Brain Mapp* 30:1060–1067. doi: 10.1002/hbm.20566
- Sepulcre J, Masdeu JC, Pastor MA, et al (2009) Brain pathways of verbal working memory. *Neuroimage* 47:773–778. doi: 10.1016/j.neuroimage.2009.04.054
- Sepulcre J, Masdeu JC, Sastre-Garriga J, et al (2008) Mapping the brain pathways of declarative verbal memory: Evidence from white matter lesions in the living human brain. *Neuroimage* 42:1237–1243. doi: 10.1016/j.neuroimage.2008.05.038
- Sullivan GM, Feinn R (2012) Using Effect Size-or Why the P Value Is Not Enough. *J Grad Med Educ* 4:279–82. doi: 10.4300/JGME-D-12-00156.1
- Tomassini V, Johansen-Berg H, Jbabdi S, et al (2012) Relating Brain Damage to Brain Plasticity in Patients With Multiple Sclerosis. *Neurorehabil Neural Repair* 26:581–593. doi: 10.1177/1545968311433208
- Tombaugh T (2006) A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). *Arch Clin Neuropsychol* 21:53–76. doi: 10.1016/j.acn.2005.07.006
- Vellinga MM, Geurts JJG, Rostrup E, et al (2009) Clinical correlations of brain lesion distribution in multiple sclerosis. *J Magn Reson Imaging* 29:768–73. doi: 10.1002/jmri.21679

Association of cerebral T2-weighted lesion load and location, vascular risk factors, and cognition in a large population-based cohort

Anna Altermatt ^{a,b}, Laura Gaetano ^{b,c}, Stefano Magon ^{b,c}, Lorena Bauer ^{b,c,e}, Regina Feurer ^e, Hans Gnahn ^e, Julia Hartmann ^f, Christian L. Seifert ^e, Holger Poppert ^{e,g}, Jens Wuerfel ^{a,b}, Ernst-Wilhelm Radue ^h, Ludwig Kappos ^e, Till Sprenger ^{c,d}

^a Department of Biomedical Engineering, University of Basel, Gewerbestrasse 14, 4123 Allschwil, Switzerland

^b Medical Image Analysis Center (MIAC) AG, Mittlere Strasse 83, 4031 Basel, Switzerland

^c Department of Neurology, University of Basel Hospital, Petersgraben 4, 4031 Basel, Switzerland

^d Department of Neurology, DKD Helios Klinik Wiesbaden, Aukammallee 33, 65191 Wiesbaden, Germany

^e Department of Neurology, Technische Universität München, Ismaningerstrasse 22, 81675 Muenchen, Germany

^f INVADE e.V., Institut für Versorgungsforschung, Karl-Böhm-Strasse 32, 85598 Baldham, Germany

^g Department of Neurology, Helios Dr. Horst Schmidt Kliniken Wiesbaden, Ludwig-Erhard-Strasse 100, 65199 Wiesbaden, Germany

^h Biomedical Research and Education GmbH, Mittlere Strasse 83, 4031 Basel, Switzerland

Corresponding author: Anna Altermatt; Email: anna.altermatt@unibas.ch; Phone: +41 61 265 43 18

Abstract

Background: Subcortical T2-weighted (T2w) lesions are very common in elderly subjects and can be associated with dementia. However, little is known about the strategic lesion distribution and how lesion patterns relate to vascular risk factors and cognitive impairment.

Aim: The aim of this study was to analyze the association between T2w lesion load and location, vascular risk factors, and cognitive impairment in a large population-based cohort of elderly subjects.

Methods: 1017 patients participating in a large prospective cohort study (INtervention project on cerebroVAscular disease and Dementia in the district of Ebersberg, INVADE) were analyzed. Cerebral T2w white matter and deep gray matter lesions, the so-called white matter hyperintensities (WMHs), were outlined semi-automatically on fluid attenuated inversion recovery images and normalized to standard stereotaxic space (MNI152) by non-linear registration. Separate lesion distributions were obtained for two protocol-defined risk groups and according to the Framingham score. Moreover, the vascular risk as assessed using the two risk ratings were mapped to specific locations using a voxel-based lesion probability mapping approach. Moreover, we assessed the relation between lesion location and cognitive impairment (demographically adjusted z-scores of the Consortium to Establish a Registry for Alzheimer's Disease Neuropsychological Assessment Battery Plus, CERAD-NAB Plus) using voxel-based statistics ($\alpha=0.05$).

Results: A total of 878 out of 1017 subjects (86%) had evaluable MRI data and were included in the analyses (mean age: 68.2 ± 7.6 years, female: 515). In the INVADE protocol-defined risk assessment, patients in the high-risk group were characterized by a significantly higher age, a higher proportion of men, and a higher lesion load ($p < 0.001$). Similarly, patients with an intermediate Framingham risk score were characterized by a higher age, male sex, and a higher lesion load ($p < 0.05$) compared to patients with a low Framingham risk score. Statistically significant associations were found for the INVADE protocol-defined or the Framingham risk groups and several of the CERAD-NAB Plus subscores ($p < 0.05$). Voxels with significant associations to the subjects' cerebrovascular risk profiles were mainly found at periventricular locations principally affecting commissural and large projection fibers and the caudate nucleus. No WMH locations were found to be significantly associated with cognitive impairment.

Conclusion: Age, gender, several cognitive scores, and WMH lesion load were shown to be significantly associated with vascular risk factors in a population of elderly, but cognitively preserved adults. Moreover, periventricular WMH lesions affecting commissural and large projection fibers and the caudate nucleus were identified to significantly relate to the subjects' cerebrovascular risk profiles.

Keywords: magnetic resonance imaging, cognition, lesion probability mapping, white matter hyperintensity, vascular risk, small vessel disease.

Introduction

Cerebral small vessel disease (SVD) is characterized by white matter T2-weighted (T2w) hyperintensities (WMHs), lacunar infarcts, and cerebral microbleeds, and has been shown to be an important etiological factor for the accrual of cognitive impairment and dementia (Rincon and Wright, 2014). WMH lesion burden, in particular, has been shown to be more common and extensive in patients with cardiovascular risk factors and symptomatic cerebrovascular disease (Abraham et al., 2016). However, little is known to date about the strategic distribution of WMHs in subcortical brain regions and how lesion patterns relate to vascular risk factors and cognitive impairment.

We here employed magnetic resonance imaging (MRI) analyses in a large population-based cohort of elderly subjects (INtervention project on cerebroVAscular disease and Dementia in the district of Ebersberg, INVADE) to better understand the association between WMH load and location, vascular risk factors, and the impact of these lesions on cognitive impairment. More in detail, patients were grouped according to their vascular risk profiles and compared in terms of demographic features, WMH load and location, and neuropsychological test performance. A lesion probability mapping approach was used to identify lesion locations which correlate with the participants' risk profiles. Finally, the same lesion mapping approach was employed to identify lesion locations, which specifically correlate with cognition, thereby trying to identify locations being specific for different cognitive domains and to shed light on the association of damage to distinct white matter (WM) tracts and deep grey matter (DGM) regions and cognitive impairment in the elderly.

Materials and Methods

2.1 Study population

Between 2001 and 2003, 3908 patients born before 1946 were recruited for the INVADE study. All patients underwent an assessment session that included the medical history, different physical examinations and the evaluation of risk factors based on distinct investigations (12-lead electrocardiography; venous blood sample; ultrasound examination of the carotid arteries; measurement of ankle-brachial-index). From April 2010 to November 2011, MRI and a neuropsychological examination (Consortium to Establish a Registry for Alzheimer's Disease Neuropsychological Assessment Battery Plus, CERAD-NAB Plus; Morris et al., 1989; Schmid et al., 2014) were additionally performed in a subgroup, termed "Neuroprevention-1" cohort, which included 1017 patients. Inclusion criteria were defined as follows: patients were born before 1946; patients were living in the district of Ebersberg, southern Bavaria (Germany); patients were able to give informed consent to study participation. Exclusion criteria were defined as follows: contraindications for MRI, severe claustrophobia, known history of stroke (ischemic stroke or intracerebral hemorrhage), refusal of getting informed about potentially relevant incidental findings on MRI, being unable to give informed consent and lack of cooperation for either the MRI- or the neuropsychological assessment (like deafness, blindness or existing movement disorders).

2.2 MR imaging protocol and lesion segmentation

All MR scans were acquired on a Philips Gyroscan NT 1.5 Tesla scanner. The MRI protocol included a highly resolving T1-weighted turbo field echo (T1w TFE, resolution = $0.96 \times 0.96 \times 1 \text{ mm}^3$, TR = 8 ms, TE = 4 ms, $\alpha = 8^\circ$), a T2*-weighted (T2*w; resolution = $0.9 \times 0.9 \times 5 \text{ mm}^3$ [interpolated to $0.45 \times 0.45 \times 5 \text{ mm}^3$], slice spacing = 1 mm, TR = 753 ms, TE = 23 ms, $\alpha = 22^\circ$) and a fluid attenuated inversion recovery sequence (FLAIR; resolution = $0.9 \times 0.9 \times 5 \text{ mm}^3$ [interpolated to $0.45 \times 0.45 \times 5 \text{ mm}^3$], slice spacing = 1 mm, TR = 6000 ms, TE = 150 ms, TI = 2000 ms). The brain was extracted automatically on the T2*w images (FMRIB's brain extraction tool, BET; using default settings, fractional intensity threshold of 0.4; Smith, 2002). Segmentation of WMHs was done at MIAC AG (Medical Image Analysis Center AG, Basel, Switzerland), blinded to clinical information. WMHs were outlined semi-automatically on the FLAIR contrast using a previously described procedure (Magon et al., 2014) with the software AMIRA 3.1.1 (Mercury Computer System Inc.), according to MIAC AG internal commercial grade standards. Lesion masks were dilated slice-by-slice by one voxel, because diffusion tensor imaging studies have shown that WM integrity is compromised immediately outside WMHs (Maillard et al., 2014).

2.3 Image registration

As the brain was outlined on T2*w images and lesions were segmented on FLAIR images, the T2*w brain images were first registered linearly to the FLAIR images (FMRIB's linear registration tool, FLIRT; using default settings, 6 degrees of freedom; Jenkinson and Smith, 2001; Jenkinson et al., 2002). Secondly, lesion masks and the above T2*w brain images were transformed to T1w coordinates using nearest-neighbor interpolation, applying the transformation parameters resulting from linearly registering the FLAIR brain images to the T1w. At this step, the skull was removed from the T1w whole-brain images using the brain masks generated from the T2*w images that were previously transformed into the T1w coordinate space. Then, a two-stages linear and non-linear registration (FMRIB's non-linear registration tool, FNIRT; settings as specified in the provided configuration file, Jacobian range chosen as 0.1 - 10; Andersson et al., 2007) was carried out to align the T1w images to the Montreal Neurological Institute standard brain template (MNI152, resolution of

2x2x2 mm³; Mazziotta et al., 2001). The obtained transformation matrices were applied to the lesion masks previously registered on the T1w image again using nearest-neighbor interpolation. The quality assessment of the registration was done visually in two steps by two trained and experienced raters. In a first step, subjects were either approved or rejected by a first rater based on the quality of the registered images. In a second step, all images were again rated independently by a second rater. In case of discrepancies, an agreement was found between the two raters (consensus reading).

2.4 Risk evaluation

Two vascular risk assessments, a study-specific and an established one, were employed in this study. First, in the INVADE study-specific cerebrovascular risk assessment, patients were assigned to either a high-risk or a low-risk group. High-risk patients were defined as patients presenting with at least one of the following criteria: Ankle brachial index ≤ 0.9 , intima media thickness ≥ 1.0 mm, stenosis of the internal carotid artery $\geq 70\%$, atrial fibrillation, transient ischemic attack, myocardial infarction, coronary artery disease, clinical manifest peripheral artery disease or Arriba-score $\geq 20\%$ (<http://www.arriba-hausarzt.de>). Second, the Framingham risk score, an established algorithm used for estimating a patient's 10-year risk of developing cardiovascular disease, was computed for each patient (D'Agostino et al., 2008).

2.5 Cognition scores

Neuropsychological testing was carried out using the CERAD-NAB Plus test battery (Morris et al., 1989; Schmid et al., 2014). CERAD-NAB Plus was originally developed to detect cognitive impairment associated with early Alzheimer Disease and has become a frequently used tool in clinical settings for cognitive impairment in general. It includes nine subtests measuring executive functions (Verbal Fluency, Trail Making Tests A and B), language (modified Boston Naming Test), verbal episodic learning (Word List Learning), verbal episodic memory (Word List Delayed Recall and Recognition), constructional praxis (Figure Learning), visual episodic memory (Figure Delayed Recall), processing speed (Trail Making Tests A and B), and frontal functions (S Words). Three additional measures were computed: the total number of intrusions at Wordlist Learning and Wordlist Delayed Recall (Wordlist Intrusions), the proportion of correct words recalled during verbal delayed recall relative to the words recalled during the learning process (Wordlist Savings), and the proportion of correctly reproduced figures at Figures Recall relative to Figures Learning (Figures Savings) (Mistridis et al., 2015). All subtest scores were transformed into z-scores, standardized for age, gender and education, using the integrated standardization program of the CERAD-NAB Plus German version.

2.6 Statistical analyses

Demographic data, WMH lesion load, and standardized CERAD-NAB Plus z-scores were compared between the study-specific risk groups and between subjects with low ($F < 10$), intermediate ($10 \leq F < 20$), and high Framingham risk ($20 \leq F < 30$) using a paired t-test for normally distributed variables and the Wilcoxon rank sum test for non-normally distributed variables. The Fisher exact test was used to compare gender differences between the risk groups.

Then, the transformed binary lesion masks of all successfully registered subjects were overlaid in MNI152 space. In this way, a probabilistic lesion distribution map was obtained. Separate lesion distribution maps were obtained for the two study-specific risk groups and subjects with low, intermediate and high Framingham scores. For the according spatial lesion distribution maps, the number of voxels where at least 10% of subjects with evaluable MRI data had a lesion was identified.

The spatial lesion mapping analyses were performed in a voxel-wise fashion. The aim was to relate the individual results of the INVADE protocol-defined risk evaluation, the Framingham risk score, and the CERAD-NAB Plus subscores

to specific lesion locations in the brain. For this purpose, two groups were defined for each voxel: one including all the patients with a lesion in that particular voxel, and one including all the patients without a lesion there (Bates et al., 2003). Then, the INVADE study-specific risk scores were compared between these two groups using a logistic regression model. The Framingham risk score and the demographically corrected z-scores of the CERAD-NAB Plus subscores were compared between these two groups using the Wilcoxon rank sum test. The same analyses were repeated simply including the subjects with particularly high/low cognition scores ($z < -1.03$ and $z > 1.03$) or simply including the subjects belonging to the study-specific high-risk group. All statistical tests were corrected for multiplicity using a Benjamini & Hochberg correction. The whole statistical analysis was implemented in R (<http://r-project.org/>). The WM fibers passing through the voxels with significant associations were identified by overlaying the map of significant voxels and the John Hopkins University (JHU) WM labels atlas (Mori and Crain, 2005) for each score. Similarly, DGM structures affected by significant relations were identified by overlaying the map of significant voxels for each score and the Montreal Neurological Institute (MNI) structural atlas (Mazziotta et al., 2001). Both atlases are available in the FMRIB Software Library (FSL).

Results

3.1 Demographic data, disease features, and cognition scores according to risk groups

A total of 878 out of 1017 subjects (86%) had evaluable MRI data and were included in the analyses. Demographic data, MRI features, and demographically adjusted z-scores of the CERAD-NAB Plus subtests of the included subjects are reported in Table 1. Overall, participants had a mean age of 68 years, a male:female ratio of roughly 1:1.7, and a WMH lesion load of 5065 mm³. Average neuropsychological test scores as measured using the CERAD-NAB Plus battery were comparable to those of a healthy population.

In the study-specific risk assessment, patients in the INVADE high-risk group were characterized by a significantly higher age ($p < 0.001$), a higher proportion of men ($p < 0.001$), and a higher lesion load ($p < 0.001$). Moreover, statistically significant differences were observed between INVADE high-risk and low-risk groups for the ‘Figure Recall’ and ‘TMT A’ subtests ($p < 0.05$). Similarly, patients with an intermediate Framingham risk were characterized by a higher age ($p < 0.001$), male sex ($p < 0.001$), and a higher lesion load ($p < 0.05$) compared to patients with a low Framingham risk score. Statistically significant associations were found for the Framingham risk groups and several cognition scores (Word List Savings/ Recognition, Figure Learning/ Recall/ Savings, Trail Making Test A) at a significance threshold of $p < 0.05$ (Table 1).

		INVADE study-specific risk				Framingham risk			
		Total (N=878)	Low-risk (N=471)	High-risk (N=407)		F < 10 (N=442)	10 ≤ F < 20 (N=313)	20 ≤ F < 30 (N=104)	
Demographics	N	Mean ± SD	Mean ± SD	Mean ± SD	p	Mean ± SD	Mean ± SD	Mean ± SD	p
Age [years]	864	68.2 ± 7.6	65.8 ± 7.4	70.9 ± 6.9	< 0.001	65.6 ± 6.9	70.8 ± 7.2	71.9 ± 7.0	< 0.001 ^{a,b}
Sex [male/female]	878	363 / 515	135 / 336	228 / 179	< 0.001	56 / 386	205 / 108	91 / 13	< 0.001 ^{a,b,c}
Education [years]	864	10.7 ± 1.9	10.8 ± 1.9	10.6 ± 1.9	-	10.8 ± 1.8	10.6 ± 2	10.5 ± 2.1	-
MRI features									
Lesion load [mm ³]	878	5065 ± 7199	3981 ± 5445	6319 ± 8642	< 0.001	4421 ± 6417	5916 ± 7734	5666 ± 8766	0.001 ^a , 0.021 ^b
CERAD-NAB Plus z-scores									
Verbal Fluency	863	-0.19 ± 0.97	-0.15 ± 1.03	-0.23 ± 0.91	-	-0.16 ± 0.99	-0.19 ± 0.98	-0.22 ± 0.85	-
mBNT	863	0.37 ± 1.05	0.39 ± 1.01	0.35 ± 1.10	-	0.39 ± 1.01	0.36 ± 1.12	0.35 ± 1.01	-
Word List Learning	864	-0.31 ± 1.20	-0.34 ± 1.20	-0.29 ± 1.21	-	-0.34 ± 1.17	-0.25 ± 1.26	-0.37 ± 1.16	-
Word List Recall	863	-0.06 ± 1.11	-0.06 ± 1.11	-0.07 ± 1.10	-	-0.08 ± 1.10	0.01 ± 1.11	-0.21 ± 1.09	-
Word List Intrusions	862	0.06 ± 1.01	0.05 ± 0.97	0.07 ± 1.05	-	0.04 ± 0.99	0.07 ± 1.03	0.11 ± 1.05	-
Word List Savings	863	-0.15 ± 1.28	-0.15 ± 1.27	-0.15 ± 1.30	-	-0.11 ± 1.33	-0.10 ± 1.24	-0.41 ± 1.20	0.034 ^b , 0.027 ^c
Word List Recognition	863	-0.13 ± 1.14	-0.13 ± 1.10	-0.12 ± 1.19	-	-0.13 ± 1.07	-0.03 ± 1.15	-0.33 ± 1.35	0.030 ^c
Figure Learning	864	0.30 ± 1.15	0.34 ± 1.04	0.26 ± 1.26	-	0.40 ± 1.08	0.22 ± 1.17	0.19 ± 1.33	0.030 ^a
Figure Recall	860	0.18 ± 1.32	0.27 ± 1.33	0.08 ± 1.31	0.041	0.30 ± 1.34	0.13 ± 1.29	-0.09 ± 1.33	0.009 ^b
Figure Savings	860	0.03 ± 1.00	0.09 ± 1.00	-0.03 ± 1.00	-	0.09 ± 0.98	0.01 ± 1.01	-0.13 ± 1.08	0.045 ^b
S Words	862	0.46 ± 1.04	0.46 ± 1.06	0.47 ± 1.01	-	0.42 ± 1.07	0.50 ± 1.00	0.47 ± 1.06	-
TMT A	860	-0.15 ± 1.16	-0.07 ± 1.19	-0.23 ± 1.11	0.048	0.00 ± 1.21	-0.27 ± 1.08	-0.33 ± 1.12	0.002 ^a , 0.014 ^b
TMT B	828	-0.17 ± 1.22	-0.15 ± 1.30	-0.20 ± 1.12	-	-0.11 ± 1.33	-0.24 ± 1.08	-0.20 ± 1.13	-
TMT B/A	827	-0.08 ± 1.13	-0.1 ± 1.18	-0.05 ± 1.07	-	-0.11 ± 1.20	-0.05 ± 1.07	-0.03 ± 0.97	-

Table 1: Characteristics of the study cohort according to risk groups. Demographic data, MRI features, and CERAD-NAB Plus z-scores were compared between the INVADE study-specific risk groups and between subjects with low ($F < 10$), intermediate ($10 \leq F < 20$), and high Framingham risk ($20 \leq F < 30$) using a paired t-test for continuous variables and a Fisher exact test for gender (a: comparison of low and intermediate Framingham risk, b: comparison of low and high Framingham risk, c: comparison of intermediate and high Framingham risk).

3.2 Spatial distribution of SVD lesions

Registered SVD binary lesion masks of all 878 participants were overlaid on the MNI brain template. Highest lesion occurrence was seen bilaterally in the periventricular WM, where lesion probability reached a maximum of 72% (Figure 1). Separate lesion distributions were obtained for the two INVADE study-specific risk groups and subjects with low, intermediate and high Framingham scores. Subjects in the protocol-defined high-risk group showed a slightly higher and more extended periventricular lesion occurrence (73%) compared to the low-risk group (71%). Moreover, the number of voxels where at least 10% of the subjects had a lesion was higher in the INVADE high-risk ($n_{10} = 8293$) versus low-risk ($n_{10} = 4467$) groups (Figure 1). Similarly, an increasing periventricular lesion load was observed from low (71%; $n_{10} = 5057$) to intermediate (74%, $n_{10} = 7625$) Framingham risk groups. For the patients in the highest Framingham risk group, a slight increase in the lesion overlay (75%) was observed, while the number of voxels affected by a lesion in at least 10% of the subjects was lower ($n_{10} = 7010$) (not shown).

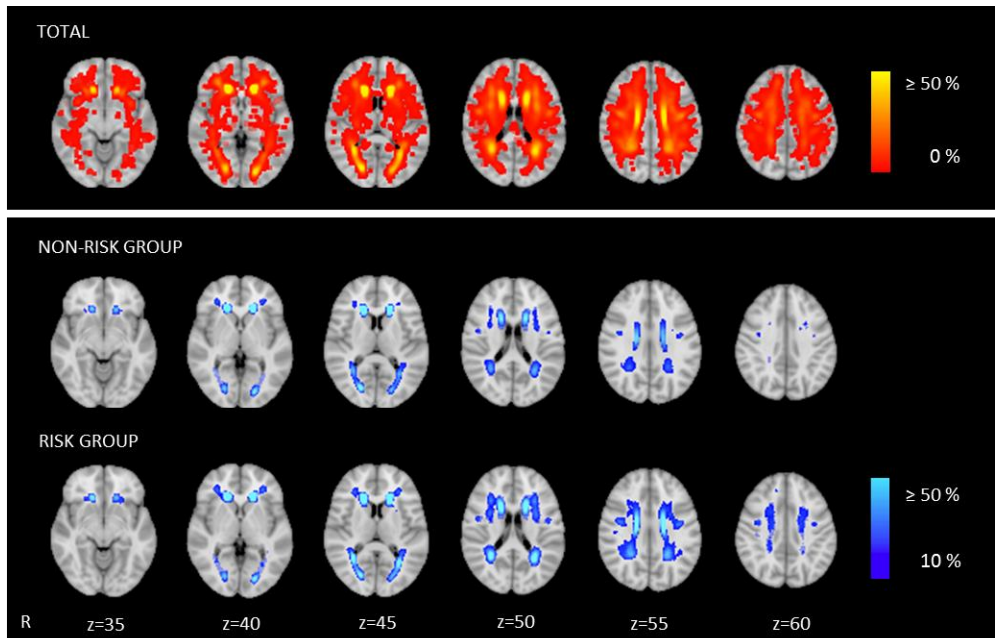


Figure 1: Lesion distributions. Sum of spatially normalized lesions of the total study cohort, per INVADE protocol-defined risk groups. The risk group-wise lesion distributions were thresholded at a minimal lesion occurrence of 10%. Only the results of the protocol-defined risk groups are shown.

3.3 Associations of lesion location and clinical scores

Lesion mapping results are shown in Figure 2, Table 2 shows the WM fiber tracts and the DGM regions affected by the significant areas. Overall, the lesion probability mapping analyses showed a similar number of voxels with significant associations to the subjects' risk profiles (INVADE protocol-defined risk assessment: $n = 2602$ voxels; Framingham risk score: $n = 2485$ voxels). Patients in the higher risk group had a more extensive lesion load mainly at periventricular locations, affecting the corpus callosum, bilaterally the superior corona radiata, and the right caudate. In patients with higher Framingham risk scores, there was stronger lesional involvement of the body of the corpus callosum, the superior corona

radiata, the right posterior corona radiata, the left thalamic radiation, and the caudate in both hemispheres compared to subjects with intermediate and low Framingham scores.

No brain lesion locations were identified to significantly impact on the patients' cognitive performance as assessed using the subscores of the CERAD-NAB Plus test battery. Moreover, also if the subjects with particularly high/low cognition scores ($z < -1.03$ and $z > 1.03$) or only the subjects belonging to the INVADE protocol-defined high-risk group were included in the analyses, no lesion locations significantly influencing the patients' neuropsychological test performance could be identified.

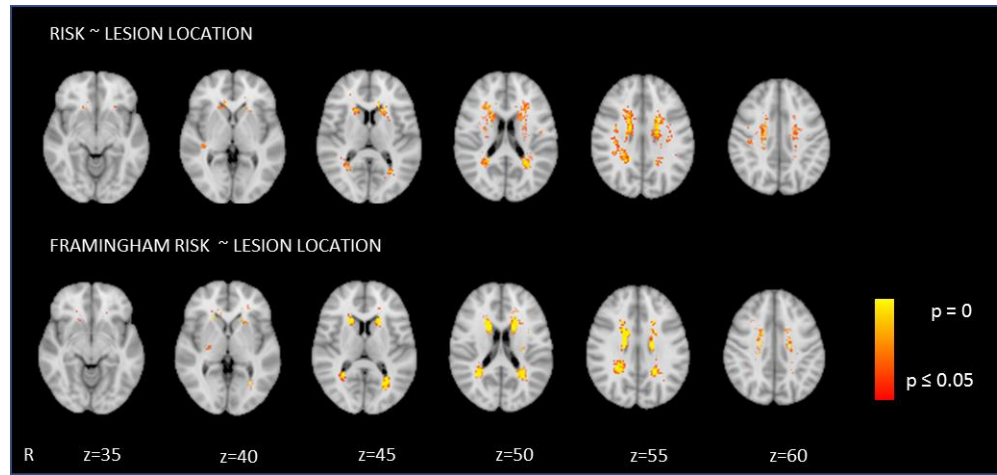


Figure 2: Significant correlations between lesion location and vascular risk. The maps depict voxels with statistically significant ($p < 0.05$) association to disability scores after applying ordinal regression (in case of INVADE study-specific risk) or the Wilcoxon rank sum test (in case of Framingham risk score) in standard stereotaxic space.

WM tracts	INVADE protocol-defined risk		Framingham risk	
	Percentage of brain region affected by lesional voxels	Percentage of voxels in brain region	Percentage of brain region affected by lesional voxels	Percentage of voxels on brain region
Body of corpus callosum	9.1	6.1	10.5	7.3
Superior corona radiata R	22.3	7.9	18.6	6.9
Superior corona radiata L	24.6	8.7	15.5	5.8
Posterior corona radiata R	26.1	4.5	31.6	5.8
Posterior thalamic radiation L	14.4	2.7	27.4	5.3
DGM region				
Caudate R	22.2	6.0	28.0	8.0
Caudate L	17.3	4.7	23.4	6.6

Table 2: Table showing the WM tracts and DGM regions where voxels with significant associations occur. Only regions where at least 5% of the significant voxels were located for one disability score are shown. The affected WM fibers were identified using the JHU WM labels atlas (Mori et al., 2005) and the involved DGM structures were determined using the MNI structural atlas (Mazziotta et al., 2001) and compared between protocol-defined and Framingham risk groups.

Discussion

We here studied the association between WMH load and location, vascular risk factors, and cognitive impairment in a large population-based cohort of 878 elderly adults from the INVADE “Neuroprevention-1” study. Patients were grouped according to their vascular risk profile and compared in terms of demographic features, WMH load and location, and neuropsychological test performance. In the INVADE protocol-defined risk assessment, patients in the high-risk group were shown to be characterized by a significantly higher age, male sex, equal or worse cognition, a higher lesion load and a more extended periventricular lesion occurrence compared to patients in the low-risk group. Voxels with significant associations to the subjects’ cerebrovascular risk profiles were mainly found in periventricular areas, while no WMH locations were found to be associated with cognitive impairment in this study.

WMH lesions, lacunar infarcts, and cerebral microbleeds are hallmarks of cerebral microangiopathy and important risk factors leading to cognitive impairment and dementia (Rincon and Wright, 2014). WMH burden in particular has been shown to be more common and extensive in patients with cardiovascular risk factors and symptomatic cerebrovascular disease (Abraham et al., 2016). For this reason, we aimed at analyzing more in depth the interplay between lesion load and location, vascular risk factors, and cognitive impairment. As expected, patients with high-risk profiles were shown to be characterized by a higher lesion load compared to patients with a lower risk. In addition to this well-established relationship, the current study adds the finding on a predilection for periventricular lesion location, more specifically at locations of commissural and large projection fibers and the caudate nucleus, in patients with a more pronounced vascular risk profile. The predilection of periventricular lesions could relate partly to the vascular anatomy of the affected WM, which has been shown to play a crucial role in lesion development in regions located at arterial watershed or border zones (Martinez and Smith, 2017).

Regarding the relation between cerebrovascular risk profiles and cognition, patients with vascular risk profiles were shown to perform worse in several cognitive domains (reflecting executive functions and processing speed, verbal episodic learning and memory, constructional praxis and visual episodic memory) compared to patients with a lower cerebrovascular risk. The lack of specific WMH locations relating to cognition in our study could be attributed to several reasons. First of all, the present patient cohort represents a healthy population of elderly adults showing normal cognition on average. As a consequence, the relative number of cognitively impaired subjects might be too small to allow for the detection of specific lesion locations being associated with a pathological performance in any of the cognitive scores. Moreover, effects of WMH on cognition could be cumulative, i.e. cognitive impairment might result from a progressive and accumulating diffuse brain damage rather than from singular lesions at well-defined locations (Wardlaw et al., 2015).

This study addresses for the first time the association between demographic features, vascular risk factors, cognitive impairment and WMH burden in a dataset of this size in the elderly. The present study is based on an elderly population with a relevant WMH burden, but still being cognitively preserved on average. Indeed, since WMHs can be measured quantitatively and non-invasively in asymptomatic patients, they are considered as potential screening markers for the identification of subjects at risk of developing clinically manifest cerebrovascular disease such as stroke or dementia.

There are some methodological limitations in our study that should be considered. First, the T2*w and FLAIR imaging sequences were characterized by a relatively large slice thickness. When registering the initial images to the isotropic MNI brain template, missing intensity values perpendicular to the slice plane had to be interpolated. We therefore cannot exclude slight inaccuracies in the normalization and registration processes. However, we consider these errors to be small as all images at all processing steps were rigorously controlled by visual inspection. Second, the imaging protocol did not include

MRI sequences such as diffusion tensor images allowing to directly identify WM tracts traveling through the regions, where associations between WMH frequency and vascular risk groups were observed.

In the future, the current approach could be extended to relate WMH patterns at baseline to progression in cognitive decline. Moreover, computer learning strategies could be used to stratify the present study population based on observed lesion patterns or cognition scores, providing a potential clinically useful prognostic tool at a single subject level (for a review, see Orrù et al., 2012).

In summary, age, gender, several cognitive scores and WMH lesion load were shown to be associated with vascular risk factors in a population of elderly adults. Periventricular lesion locations affecting commissural and large projection fibers and the caudate nucleus were identified to significantly relate to the subjects' cerebrovascular risk profiles.

References

- Abraham, H.M.A., Wolfson, L., Moscufo, N., Guttmann, C.R.G., Kaplan, R.F., White, W.B., 2016. Cardiovascular risk factors and small vessel disease of the brain: Blood pressure, white matter lesions, and functional decline in older persons. *J. Cereb. Blood Flow Metab.* 36, 132–142. <https://doi.org/10.1038/jcbfm.2015.121>
- Andersson, J., Jenkinson, M., Smith, S., 2007. Non-linear registration aka s spatial normalisation FMRIB Technial Report TR07JA2.
- Bates, E., Wilson, S.M., Saygin, A.P., Dick, F., Sereno, M.I., Knight, R.T., Dronkers, N.F., 2003. Voxel-based lesion-symptom mapping. *Nat. Neurosci.* 6, 448–50. <https://doi.org/10.1038/nn1050>
- D’Agostino, R.B., Vasan, R.S., Pencina, M.J., Wolf, P.A., Cobain, M., Massaro, J.M., Kannel, W.B., 2008. General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study. *Circulation* 117, 743–753. <https://doi.org/10.1161/CIRCULATIONAHA.107.699579>
- Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–41. <https://doi.org/10.1006/nimg.2002.1132>
- Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–56. [https://doi.org/10.1016/S1361-8415\(01\)00036-6](https://doi.org/10.1016/S1361-8415(01)00036-6)
- Magon, S., Gaetano, L., Chakravarty, M.M., Lerch, J.P., Naegelin, Y., Stippich, C., Kappos, L., Radue, E.-W., Sprenger, T., 2014. White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: a longitudinal study. *BMC Neurosci.* 15, 106. <https://doi.org/10.1186/1471-2202-15-106>
- Maillard, P., Fletcher, E., Lockhart, S.N., Roach, A.E., Reed, B., Mungas, D., DeCarli, C., Carmichael, O.T., 2014. White Matter Hyperintensities and Their Penumbra Lie Along a Continuum of Injury in the Aging Brain. *Stroke* 45, 1721–1726. <https://doi.org/10.1161/STROKEAHA.113.004084>
- Martinez Sosa, S., Smith, K.J., 2017. Understanding a role for hypoxia in lesion formation and location in the deep and periventricular white matter in small vessel disease and multiple sclerosis. *Clin. Sci.* 131, 2503–2524. <https://doi.org/10.1042/CS20170981>
- Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Boomsma, D., Cannon, T., Kawashima, R., Mazoyer, B., 2001. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 356, 1293–322. <https://doi.org/10.1098/rstb.2001.0915>
- Mistridis, P., Krumm, S., Monsch, A.U., Berres, M., Taylor, K.I., 2015. The 12 Years Preceding Mild Cognitive Impairment Due to Alzheimer’s Disease: The Temporal Emergence of Cognitive Decline. *J. Alzheimers. Dis.* 48, 1095–107. <https://doi.org/10.3233/JAD-150137>
- Mori, S., Crain, B.J., 2005. *MRI Atlas of Human White Matter*. Elsevier, Amsterdam.
- Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum, G., Mellits, E.D., Clark, C., 1989. The Consortium to Establish a Registry for Alzheimer’s Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer’s disease. *Neurology* 39, 1159–65. <https://doi.org/10.1212/WNL.39.9.1159>
- Orrù, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G., Mechelli, A., 2012. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neurosci. Biobehav. Rev.* 36, 1140–1152. <https://doi.org/10.1016/j.neubiorev.2012.01.004>

- Rincon, F., Wright, C.B., 2014. Current pathophysiological concepts in cerebral small vessel disease. *Front. Aging Neurosci.* 6, 24. <https://doi.org/10.3389/fnagi.2014.00024>
- Schmid, N.S., Ehrensperger, M.M., Berres, M., Beck, I.R., Monsch, A.U., 2014. The Extension of the German CERAD Neuropsychological Assessment Battery with Tests Assessing Subcortical, Executive and Frontal Functions Improves Accuracy in Dementia Diagnosis. *Dement. Geriatr. Cogn. Dis. Extra* 4, 322–334. <https://doi.org/10.1159/000357774>
- Smith, S.M., 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155. <https://doi.org/10.1002/hbm.10062>
- Wardlaw, J.M., Valdés Hernández, M.C., Muñoz-Maniega, S., 2015. What are white matter hyperintensities made of? Relevance to vascular cognitive impairment. *J. Am. Heart Assoc.* 4, 1140. <https://doi.org/10.1161/JAHA.114.001140>

5 Relevance of segmentation validation in brain atrophy measurements

5.1 Clinico-radiological association

Studies have shown that brain atrophy is an important marker of MS or SVD disability status and progression, showing a better clinico-radiological association than conventional lesion assessments. Yet, one cause for concern relates to the repeatability of individual measurements. Section 5.1 introduces the role of WM and GM segmentation, and the factors which potentially confound computed volumes. Then, different methods for segmentation validation are discussed, including manual expert delineation, digital and physical phantoms (Section 5.2). Finally, existing anatomical physical phantoms are presented together with their most relevant limitations (Section 5.3).

Clinico-radiological association: Brain atrophy is an MRI measure reflecting brain volume loss through neurodegeneration. In MS, studies linking brain atrophy to clinical impairment state that irreversible tissue destruction – to a greater extent than can be explained by conventional lesion assessments – is an important determinant of disability and disease progression [54]. Indeed, whole-brain atrophy was shown to have a stronger imaging association with physical disability and cognitive impairment, and to be a stronger predictor of future disability than total lesion burden [55]. Moreover, regional brain volume loss (e.g. GM atrophy [56]) may even reveal a more pronounced relation to clinical disease symptoms, and we rely on both, robust whole-brain and regional volume measures [5]. In comparison to MS, there are less data on brain atrophy in SVD [6]. Still, SVD-related atrophic changes have been shown to correlate with SVD disease progression, and GM atrophy in particular, was shown to be the major contributor to whole brain atrophy in symptomatic cerebral SVD [6, 57].

Segmentation methods: Clinically, brain atrophy can be identified from qualitative images by the detection of an increase in CSF spaces or a reduction in size of the parenchyma compared with the normal appearance for age. Such relatively simple analyses would be easy to implement in routine patient care. However, because of the limited reproducibility and precision of visually based atrophy measurements, we rely on automated or semiautomated quantitative methods to segment whole or regional brain volumes [54]. Classically, image segmentation is defined as the partitioning of an image into nonoverlapping parts that are homogeneous with respect to some characteristic such as intensity or texture [58].

In the absence of artefacts, segmentation of the main brain tissues could in principle be done by simple thresholding strategies. Indeed, the histogram of T1-weighted images of the adult brain MRI has three main peaks corresponding to WM, GM, and CSF. Images acquired in clinical or research routine are, however, noisy and deployed algorithms need to be able to cope with different types of artefacts (i.e. intensity inhomogeneity, noise, partial volume, as well as overlap in intensities of brain and nonbrain tissue) [58]. Hence, during the past decades, a wealth of dedicated segmentation algorithms in the field of medical image analyses have emerged.

Different automated segmentation tools are implemented in the most popular software packages for brain image processing, including FMRIB software library (FSL; <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki>), Statistical Parametric Mapping (SPM; www.fil.ion.ucl.ac.uk/spm), and FreeSurfer (<https://surfer.nmr.mgh.harvard.edu>). In these software packages, the algorithms underlying the standard segmentation tools all employ unsupervised intensity-based classification, which partitions an image into regions of voxels with similar intensities without using training images [59–61]. Various additional improvement are further encoded in these methods such as borrowing of local spatial information from neighboring voxels (e.g. using Markov random fields as in FSL FAST and FreeSurfer; [59, 60] or inclusion of *a priori* atlas information (e.g. atlas priors as in SPM and FreeSurfer; [60, 61]).

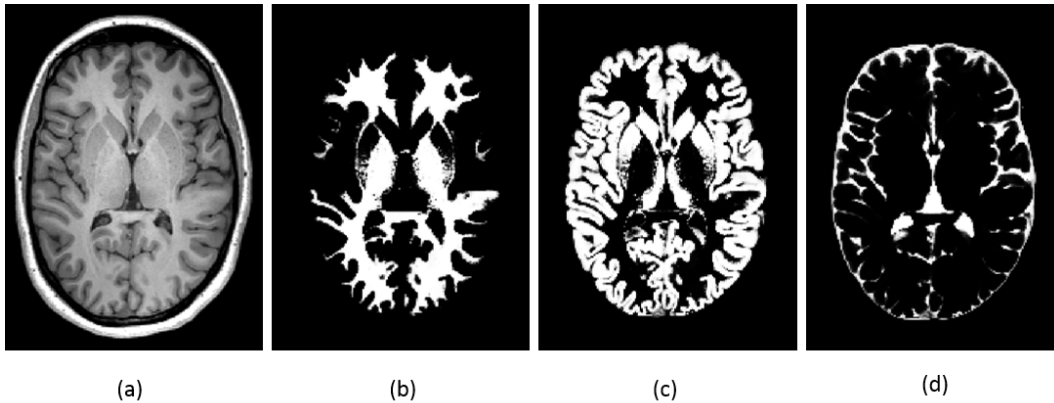


Fig. 4: Segmentation of main brain tissues. Figure depicting the segmentation of a (a) T1-weighted brain MRI into (b) WM, (c) GM, and (d) CSF. This exemplary segmentation was done using SPM [61].

Factors contributing to moderate association: Several issues linked to the quantification of brain tissues are yet thought to weaken or distort the clinico-radiological associations. The most relevant cause for concern relates to the repeatability (i.e. test-retest reliability) of individual measurements. Indeed, there are different confounding factors which can affect computed brain volumes independent of real tissue destruction. These factors can either be of technical or biological origin. Confounding technical factors are for example linked to the

patient positioning, the combination of scans from different hardware, the use of different pulse sequences, or the employment of different segmentation software. Biological factors are linked to sources of variability intrinsic to the patient, for example due to hydration levels or the menstrual cycle [62]. The influence of these factors might prevent robust comparison of single patient measurements against those in a database of normal subjects. It can further confound the analysis of longitudinal measurements from an individual, since it is crucial to know whether any apparent trend is simply due to measurement variation or a true disease effect [62].

5.2 Relevance of segmentation validation in brain atrophy measurements

To guarantee robust measurements of brain atrophy and to reliably associate clinical and radiological disease manifestations, segmentation validation requires a “ground truth” or gold standard to which the segmentation outcome can be compared [58].

Manual segmentation: Usually, the gold standard for real patient data is made by one or more expert physicians who manually outline the anatomical structures of interest. Since this is the only method to validate the real patient MRI data, this method is used very frequently in brain image analyses. A popular repository with real MRI data and available manual expert segmentations is the Internet Brain Segmentation Repository (IBSR; <https://nitrc.org/projects/ibsr/>). Despite its popularity, this validation method should be used with caution, since manual segmentations are prone to errors, highly subjective, difficult to reproduce (even by the same expert), and time-consuming (especially for large brain regions such as WM or GM) [58].

Digital phantoms: Another popular validation method is the use of software simulations. In these simulations, artificial MR images are generated digitally through a simulation of the real acquisition process. In this way, the “ground truth” is known in advance and the influence of different acquisition parameters and imaging artifacts on the segmentation outcome can be controlled. This type of validation is very flexible and efficient. The most popular simulated images used for brain MRI segmentation methods are found in the BrainWeb repository (brainweb.bic.mni.mcgill.ca/brainweb/). Another example is the Alfano digital phantom [63]. A drawback of this type of validation is that software simulations cannot take into account all factors linked to the image acquisition [58]. Indeed, processes that are optimized based on numerical simulations may fail or introduce systematic bias when presented with real scanner data if specific factors are ignored or overlooked. Moreover, although many elements linked to the image acquisition can be identified and incorporated in the signal model, yet there are some arbitrary experimental factors that remain mathematically intractable in the framework of digital simulations [64].

Physical phantoms: Since software simulations have certain limitations, segmentation validation can be done using physical phantoms, which mimic certain *in vivo* properties.

The design and construction of such physical phantoms offer less flexibility and are more labor intensive compared to digital phantoms. However, their biggest advantage relates to the fact that images of these phantoms can be acquired using the MRI hardware [58].

5.3 Anthropomorphic physical phantoms

Physical phantoms: In general, MRI phantoms are fluid-filled objects that mimic body dimensions or shape, and most are cylindrical or spherical in shape (Figure 5) [65]. There are only a few standard MRI phantoms that are routinely used for quality assurance testing. Examples are the American College of Radiology Accreditation (ACR; [66]), the Alzheimer’s Disease Neuroimaging Initiative (ADNI; Figure 5; [67]A), or the National Institute of Standards and Technology (NIST; [68]) phantoms. Devices mimicking the tissue-specific longitudinal and transverse relaxation times are typically made from a rigid container, filled with an aqueous solution doped with paramagnetic ions such as GdCl_3 , MnCl_2 , CuSO_4 , or NiCl_2 . Typically, a gelling agent (e.g. gelatin, polyacrylamide, polyvinyl alcohol, agarose, carrageenan, or agar) is added to the aqueous solution to avoid flow artifacts [65]. Still, to our knowledge, there is no commercially available phantom mimicking relaxation times and structure of the main brain compartments, and which is routinely used for segmentation validation.

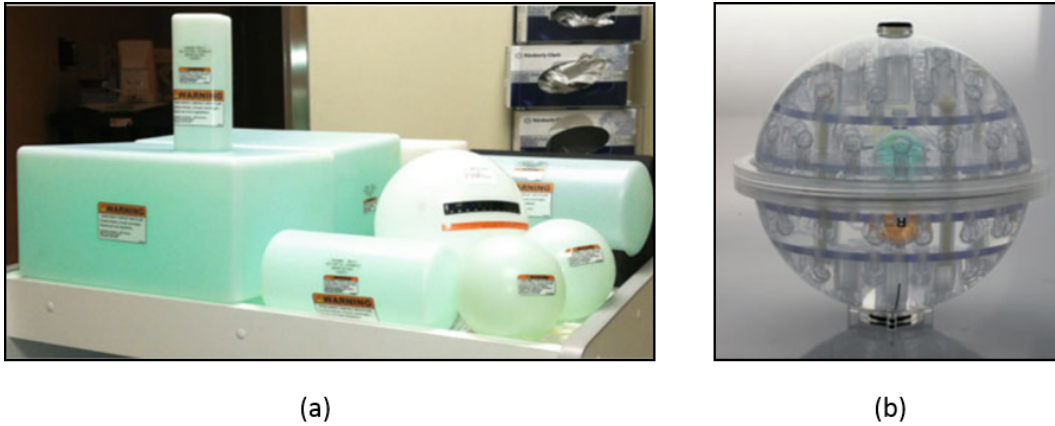


Fig. 5: Conventional physical phantoms. (a) Photograph showing several MRI loading phantoms. Each vendor provides coil-specific loading phantoms. (b) Photograph of the Magphan® Quantitative Imaging Phantom from the Alzheimer’s Disease Neuroimaging Initiative (ADNI). It consists of 165 spherical objects inside a water-filled urethane shell. The spherical objects are polycarbonate shells filled with various concentrations of CuSO_4 (figure taken from [65]).

Anthropomorphic physical phantoms: A number of physical phantoms – which are *not* commercially available – mimicking brain relaxation times and showing increased anatomical similarity to the real counterpart have been developed by research groups or academic centers (Table 3; Figure 6). Phantoms mimicking the outer brain contours (i.e. ventricle shape and cortical folding patterns) have been proposed by Chen et al. (2012) [25] and Khan et al. (2012) [26]. Curtis et al. (2010) [23] suggested the first phantom mimicking relaxation times of two brain compartments (i.e. WM and GM). In the proposed constructing procedure, 14 brain slices were separately milled from plastic, filled with the tissue-mimicking solutions, and assembled to a whole brain phantom. Kuss et al. (2011) [24] developed a head phantom consisting of a skull, an adipose layer for the skin, and a deformable brain including WM and GM compartments. Surfaces of WM and GM were mimicked by manually cutting the gels to resemble the cortical folding patterns. Another multipurpose phantom was proposed by Fujimoto et al. (2015) [27] who printed chambers from polylactic acid for WM, GM, and CSF and filled them with three different solutions. Very recently, Gopalan et al. (2017) [28] constructed a 3D-printed, anatomy-mimicking, quantitative MRI phantom for validating methods that rely on spatio-temporal modeling. Similar to the previous phantom, 3D-printed polylactic acid chambers were filled with the corresponding solutions mimicking WM, GM, and CSF.

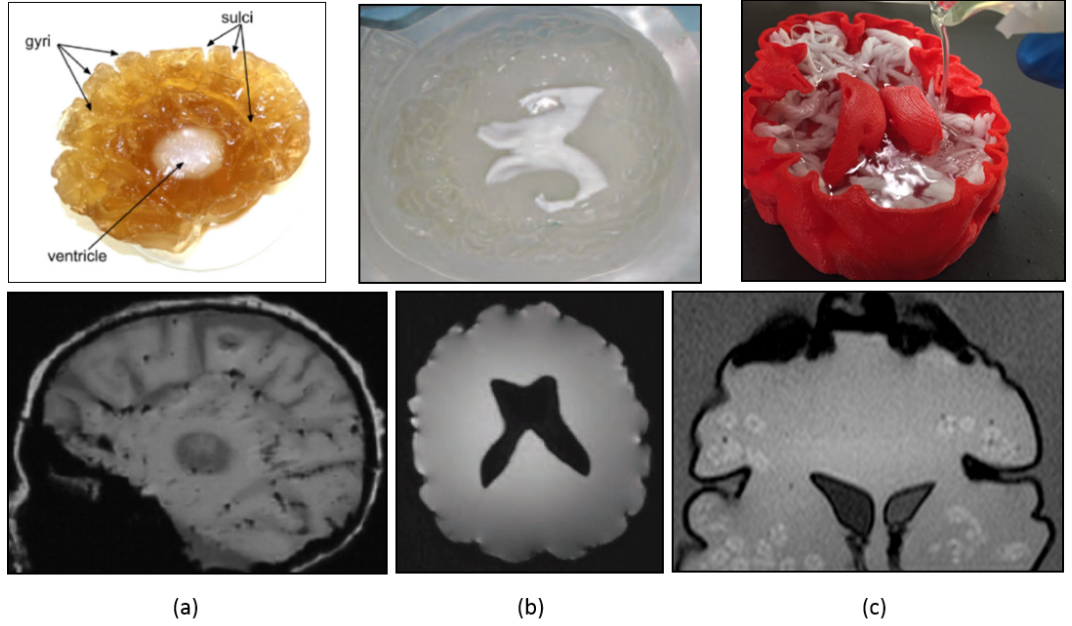


Fig. 6: Existing anthropomorphic brain phantoms. Photographs (top row) and T1-weighted MRI (bottom row) of some existing anatomical phantoms. (a) Phantom by Kuss et al. [24]; (b) Phantom by Khan et al. [26]; Phantom by Gopalan et al. [28]. These phantoms are described in detail in Table 3.

Study	Phantom compartments	Support between WM and GM compartments	Gelling and MR contrast agents	MR properties	Constructing procedure
Curtis et al., (2010), [23]	WM, GM	Slices milled from plastic	Agarose, GdCl ₃	T1, T2	Slice-by-slice
Kuss et al., (2011), [24]	WM, GM, CSF	None	Gelatin, CuSO ₄	T1, T2	Shell-wise, manually cutting shapes into gels
Chen et al., (2012), [25]	Brain parenchyma	-	Polyvinyl alcohol, CuSO ₄	T1, T2	Filling of flexible rubber mold, immersion of rubber ventricle
Khan et al., (2012), [26]	Brain parenchyma	-	Agar, Gd-DTPA	T1	Filling of robust plastic mold, immersion of plastic ventricle
Fujimoto et al., (2015), [27]	WM, GM, CSF	3D-printed chambers	Agar, MnCl ₂	T1	Filling of 3D-printed chambers
Gopalan et al., (2017), [28]	WM, GM, CSF	3D-printed chambers	Agar, NiCl ₂ or MnCl ₂	T1, T2	Filling of 3D-printed chambers

Tab. 3: Existing physical phantoms. Table describing the existing physical phantoms, which mimic T1 relaxation times of main brain tissue types and show increased anatomical similarity with real counterpart.

These existing physical phantoms showing increased anatomical similarity to the brain were hampered by a few limitations: precise T1 (or T2) relaxation times of WM and GM were only matched in a few phantoms [23, 27, 28]; contours of brain or brain regions were substantially simplified [23–28]; and some phantoms were characterized by the presence of thicker physical walls separating the solution compartments [23, 27, 28]. These limitations make it difficult/impossible to realistically mimic segmentation with these existing phantoms. For this reason, to our knowledge, no study has ever been conducted using an anthropomorphic physical phantom to validate segmentation of WM and GM in MRI.

6 An anthropomorphic physical phantom for segmentation validation

Aiming at constructing a physical brain phantom, which mimics structure and T1 relaxation properties of WM and GM, different materials and assembling procedures were elaborated and tested. An innovative framework for the construction of anatomical physical brain phantoms was established. The constructed phantom prototype is presented in the first publication presented in this chapter. Briefly, WM and GM relaxation times were reproduced using an agar gel doped with MR (and CT) contrast agent. Folding patterns of these two phantom compartments were simulated through a molding-casting procedure using 3D-printed casts and flexible silicone molds. The second publication presented in this chapter is a proof-of-principle study, where the designed phantom was used to validate automatic brain tissue segmentation. Ground truth volumes of the phantom are derived from highly-resolved CT scans acquired together with T1-weighted MR images for automatic segmentation by two popular software tools.

Publication: *Design and construction of an innovative brain phantom prototype for magnetic resonance Imaging.* Anna Altermatt, Francesco Santini, Xeni Deligianni, Stefano Magon, Till Sprenger, Ludwig Kappos, Philippe Cattin, Jens Wuerfel, Laura Gaetano. This publication has been submitted to the journal Magnetic Resonance in Medicine and is currently under review.

Publication: *Evaluation of Automated Brain MR Image Segmentation Methods: A Physical Phantom Study.* Anna Altermatt, Francesco Santini, Xeni Deligianni, Stefano Magon, Till Sprenger, Ludwig Kappos, Philippe Cattin, Jens Wuerfel, Laura Gaetano. This publication has been submitted to the journal NeuroImage and is currently under review.

Design and construction of an innovative brain phantom prototype for magnetic resonance imaging

Anna Altermatt ^{a,b}, Francesco Santini ^{a,d}, Xeni Deligianni ^{a,d}, Stefano Magon ^{b,c}, Till Sprenger ^e, Ludwig Kappos ^{a,c},
Philippe Cattin ^a, Jens Wuerfel ^{a,b}, Laura Gaetano ^{b,c}

Affiliations:

^a Department of Biomedical Engineering, University of Basel, Switzerland

^b Medical Image Analysis Center (MIAC) AG, Basel, Switzerland

^c Neurologic Clinic and Policlinic, University Hospital of Basel, Switzerland

^d Division of Radiological Physics, Department of Radiology, University Hospital of Basel, Switzerland

^e Department of Neurology, DKD HELIOS Klinik Wiesbaden, Germany

Corresponding author:

Laura Gaetano

Medical Image Analysis Center (MIAC) AG

Mittlere Strasse 83

4031 Basel

Switzerland

Mail: laura.gaetano@miac.ch

Phone: +41 61 265 43 18

Abstract

Purpose: The purpose of this project was to construct a physical brain phantom for MRI, mimicking structure and T1 relaxation properties of white matter (WM) and grey matter (GM).

Methods: The phantom design comprised two compartments, one resembling the WM and one the GM. Their T1 relaxation times, as assessed using an inversion recovery turbo spin echo sequence, were reproduced using an agar gel doped with contrast agent (CA) and their folding patterns were simulated through a molding-casting procedure using 3D-printed casts and flexible silicone molds. Three versions of the assembling procedure were adopted to build: Phantom1 without any separation; Phantom2 with a varnish layer; and Phantom3 with a thin wax layer between the compartments.

Results: Phantom1 was characterized by a pronounced diffusion of CA between the two compartments. Phantom2 and Phantom3, instead, showed relaxation times and shape comparable with the target ones identified in a healthy control subject (WM: 754 ± 40 ms; GM: 1277 ± 96 ms). Moreover, both compartments revealed intact gyri and sulci. However, the diffusion of CA made Phantom2 stable only for a short period of time. Phantom3 showed stability within a longer time window but the wax layer between the WM and GM was visible in the MRI.

Conclusion: Structural and intensity properties of the constructed phantoms are useful in evaluating and validating steps from image acquisition to image processing. Moreover, the described constructing procedure and its modular design make it adjustable to a variety of applications.

Keywords: Magnetic resonance imaging, physical phantom, brain anatomy, 3D printing, T1 relaxation time

Introduction

Phantoms have been used as tools for any type of MRI validation. They are physical or digitally-simulated objects of known geometry and/or composition with a reduced number of experimental confounds with respect to living subjects. Physical phantoms are commonly used when calibrating scanner hardware or designing pulse sequences. Digital phantoms instead, are especially prevalent in the image processing and reconstruction fields, as they facilitate quantitative comparison between algorithms and enable optimization over a wide variety of imaging sequences. While digital phantoms allow for modeling complex structures, physical phantoms are limited by the manageability of the used materials and the corresponding physical properties (1). Consequently, most existing physical phantoms are characterized by a simple geometry, i.e. spherical or cylindrical (2). Conversely, using digital phantoms, factors linked to the image acquisition can only be simulated and the testing procedure largely depends on the accurateness of this simulation (1).

In the specific context of brain MRI, a limited number of physical phantoms with increased similarity to the real counterpart, both in terms of tissue intensities and shape, have been developed in the past years (3, 4, 5, 6, 7, 8). However, some of those just mimicked one tissue (7, 8), some required the presence of walls separating different compartments to maintain the shape (5, 6), and some others produced non-realistic or substantially simplified contours of brain or brain regions (4, 5, 6, 7, 8).

To overcome those limitations, we propose an innovative design for the construction of a 3D brain phantom that, at the same time, resembles the anatomy of the main brain tissue types (i.e. WM and GM) and produces T1 relaxation times corresponding to those observed in vivo. To reach those goals, we evaluated three slightly different designs and we characterized the final properties of the phantoms, in terms of shape, intensity, and stability over time.

Methods

The phantom structures and properties were based on a healthy volunteer (26-year-old female without history of neurological and psychiatric diseases) that provided written informed consent and the study was conducted in accordance with the Declaration of Helsinki. All images were acquired at a 3 Tesla (T) scanner (Magnetom Prisma, Siemens Healthineers, Erlangen, Germany) with a 20-channel head coil.

2.1. Phantom intensity

To determine the reference values, T1 mapping was performed on the healthy brain using an inversion recovery turbo spin echo sequence (2D IR-TSE, axial slice plane, resolution = 1.3x1.3 mm², slice thickness = 5 mm, TR = 6000 ms, TE = 12 ms, TI = 30/50/100/200/500/750/1000/2000 ms, Echo train length = 7, $\alpha = 180^\circ$, total scan time = 12 min 48 s). T1 values of 2 regions of interest (ROIs) of dimensions 3x3 mm² in the white matter (WM) and 2 ROIs of dimensions 3x3 mm² in the gray matter (GM) were computed by nonlinear fitting of the average intensity inside the ROIs with the standard inversion recovery signal model (9).

To match those reference values, six different concentrations of MRI contrast agent MnCl₂ (Sigma-Aldrich Chemie GmbH, Steinheim, Germany) were diluted in 0.6 % agar solution (Morga AG, Ebnat-Kappel, Switzerland) (Supporting Table 1), that was always done in batches of 0.5 L or 1 L to ensure the reproducibility of T1 values. For each concentration, three samples were prepared in 50-mL plastic tubes (diameter equal to 28 mm) with the following procedure. Half liter of water was heated to 100 °C, 10 g of agar were added to the boiling water and the solution was kept at boiling temperature for 2 minutes 30 seconds while stirred constantly. The boiling agar solution was then poured into the prepared tubes with the different MnCl₂ volumes. Samples were mixed to ensure a homogeneous mixture of contrast agent and agar powder, and then left to cool down to room temperature in order to gel.

The samples were subsequently scanned with the same quantitative sequence (2D IR-TSE, axial slice plane, resolution = 1x1 mm², slice thickness = 5 mm, TR = 6000 ms, TE = 12 ms, TI = 30/50/100/200/500/750/1000/2000 ms, Echo train length = 7, $\alpha = 180^\circ$) and T1 of each was calculated using a 3x3 mm² ROI located in the center of the tubes to avoid edge effects. The T1 values were then used to find the quantitative relation between T1 values and contrast agent concentration by means of the relaxivity equation:

$$R_1 = \frac{1}{T_1} = \frac{1}{T_1^0} + r_1[CA], [1]$$

where R_1 is the relaxation rate constant, T_1^0 is the relaxation time of the tissue in the absence of contrast agent, r_1 is the relaxivity and $[CA]$ is the concentration of contrast agent.

Thus, MnCl₂ concentrations yielding T1 relaxation times in the ranges of WM and GM were identified. To ensure the correctness of the concentrations just selected, six control samples (three for WM and three for GM) were prepared with the previous procedure and their T1 values were estimated with the 2D IR-TSE sequence.

2.2. Phantom shape

A whole-brain isotropic T1-weighted image (3D MPRAGE, sagittal orientation, resolution = $1 \times 1 \times 1 \text{ mm}^3$, TR = 1570 ms, TI = 900 ms, TE = 2.48 ms, $\alpha = 8^\circ$, scan time = 7 min 38 s) was acquired from the healthy subject. Surfaces of WM and GM of the left hemisphere were segmented using FreeSurfer 6.0 (Harvard University, Cambridge, Massachusetts, United States). 3D Slicer (Surgical Planning Laboratory, Brigham and Women's Hospital, Boston, Massachusetts, United States) was used to extract the inner skull mesh of the hemisphere by means of the semi-automated threshold effect and change island effect modules. These surfaces, i.e. the WM, GM and skull, were converted to triangular meshes and saved as stereolithography files (Meshmixer 2.9, Autodesk, Inc., San Rafael, California, USA).

The mesh density was reduced to 50% of the initial trial count and the surfaces were smoothed to reduce noise using a smoothing filter at the lowest setting. Then, to facilitate the phantom assembling, the cortical thickness was increased by 20 mm through scaling down the WM compartment. The final surfaces were printed on a fifth-generation MakerBot Replicator+ 3D printer (MakerBot® Industries, Brooklyn, New York, United States) using polylactic acid (PLA) thermoplastic (WM: resolution = 77974 vertices, print time = 14 h 16 min, with support; GM: resolution = 78077 vertices, print time = 20 h 59 min; skull: resolution = 48558 vertices, print time = 29 h 22 min; 0.2 mm slice thickness and including support for all prints), and brushed on with a platinum-cure silicone rubber in 3 layers (shore A and hardness 30, DWR Plastics, Broadway, Newport, Shropshire, UK) to create three flexible negative molds (10), easily allowing the detachment of the paramagnetic gels in the next stages.

A multi-step procedure was adopted for the final assembly of the brain-mimicking gels: the WM sample was injected into the WM surface mold; once the sample gelled, the mold was removed; the GM mold was positioned on top; the GM sample was then injected and the corresponding mold was pulled off when the gel became solid; finally, a silicone shell mimicking the skull was positioned on top of the gel composite (Fig. 1). Three versions of this procedure were tested: no layer was put between the WM and GM gels (Phantom1), a varnish layer (MOTIP DUPLI GmbH, Haßmersheim, Germany) was sprayed on the WM gel (Phantom2), or a thin wax layer (GLOREX GmbH, Rheinfelden, Germany) was applied between the two compartments (Phantom3).

2.3. Phantom characterization

After the phantoms were completed, their final structure and properties were examined (Day1). The structural accuracy was verified using T1-weighted imaging (3D MPRAGE, sagittal slice plane, resolution = $0.82 \times 0.82 \times 1.00 \text{ mm}^3$, TR = 1570 ms, TI = 900 ms, TE = 1.34 ms, $\alpha = 8^\circ$, scan time = 6 min 38 s). To determine the intensities of WM and GM compartments, T1 mapping was performed on the phantoms using the IR-TSE sequence described in the previous setup (2D IR-TSE, horizontal slice plane, resolution = $1 \times 1 \text{ mm}^2$, slice thickness = 5 mm, TR = 6000 ms, TE = 12 ms, TI = 30/50/100/200/500/750/1000/2000 ms, Echo train length = 7, $\alpha = 180^\circ$, total scan time = 13 min 52 s). In this case, two different slices were acquired and, for every slice, average T1 intensities within two ROIs for both compartments ($3 \times 3 \text{ mm}^2$, one in the frontal area and one in the posterior part) were computed.

2.4. Stability

To assess the stability of the phantoms in terms of T1 relaxation times and shape, the same protocol (i.e. 3D MPRAGE and 2D IR-TSE) was acquired after 2, 4, 8, 15, 21, and 28 days (Day2, Day4, Day8, Day15, Day 21, Day28, respectively) after the preparation.

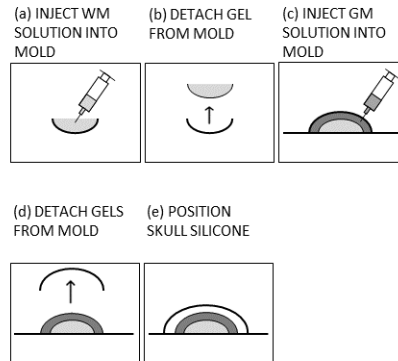


Fig. 1: Phantom Assembling. Multi-step procedure followed for the phantom assembling. Briefly, the WM gel was injected into the WM surface mold (a); once the gel set, the mold was removed (b); then, the GM mold was positioned on top and the GM gel was injected (c); after 2 hours, the GM gel became solid and the corresponding mold was pulled off (d); finally, the silicone shell mimicking the skull was positioned on top of the gel composite (e). Three versions of this procedure were tested: no layer was put between the WM and GM gels (Phantom1), a varnish layer was sprayed on the WM gel (Phantom2), or a thin wax layer was applied between the two compartments (Phantom3).

Results

The targeted values for WM and GM derived from the template brain were 754 ± 40 ms and 1277 ± 96 ms, respectively. Those values could be reproduced with MnCl_2 concentrations of 0.12 mM for the WM and 0.04 mM for the GM, leading to T1 values of 732 ± 5 ms and 1303 ± 8 ms in the WM and GM control samples, respectively (Fig. 2). Silicone molds of WM, GM, and skull were created from the 3D-printed surface meshes (Fig. 3). Using the procedure depicted in Figure 1, the WM and GM gels (Fig. 3d) were assembled and scanned.

Phantom1 was characterized by an indistinguishable intensity difference between WM and GM at Day1 (data not shown). The absence of a layer between the WM gel and the GM one allowed the MnCl_2 contrast agent to diffuse from one compartment to the other as soon as the two gels were put into contact. The application of some kind of a hydrophobic layer in Phantom2 and Phantom3, instead, prevented this process and allowed us to evaluate the entire constructing procedure.

WM and GM T1 relaxation times of Phantom2 were shown to be comparable to the values identified in the healthy control subject (Fig. 5). Both compartments were characterized by intact gyri and sulci typical of a human brain (Fig. 3e; Fig. 4), showing that the assembly process did not alter the target shape. The geometric fidelity was assessed qualitatively comparing the external surfaces (Fig. 3) as well as in-plane slices (Supporting Fig. S2) of the simplified model (i.e. the one used for the 3D printer) and the phantom. Only few issues were identified in the shape of the phantom: a very limited number of gyri showed an unusual profile because the mold was not entirely filled by the MnCl_2 -agar solution (Fig. 4, green arrow) and some small air cavities were observed at the WM/GM interface (Fig. 4, red arrows). The shape and T1 relaxation times of the GM of Phantom2 remained stable within an observed time window of 84 hours from preparation. However, due to an increase in WM T1 relaxation times, an overall decrease in the contrast difference between WM and GM was observed (Fig. 5).

Similar to Phantom2, Phantom3 showed relaxation times and shape comparable with the target ones. In this case, the diffusion of MnCl_2 was prevented by the wax layer for a longer period than in Phantom2, but it started occurring after 8 days (Fig. 5 and Supporting Fig. S3). Moreover, the applied wax layer was clearly distinguishable in the MRI (Fig. 5).

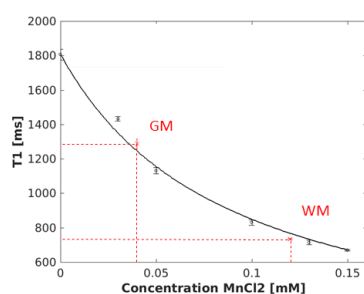


Fig. 2: Calibration curve. Six different MRI contrast agent (i.e. MnCl_2) concentrations were chosen and three samples were prepared for each concentration. The resulting T1 relaxation times were used to fit a calibration curve (line). According to this formula, the concentrations of MnCl_2 for WM and GM similar to the template brain were derived and corresponding control samples were prepared (shown in red).

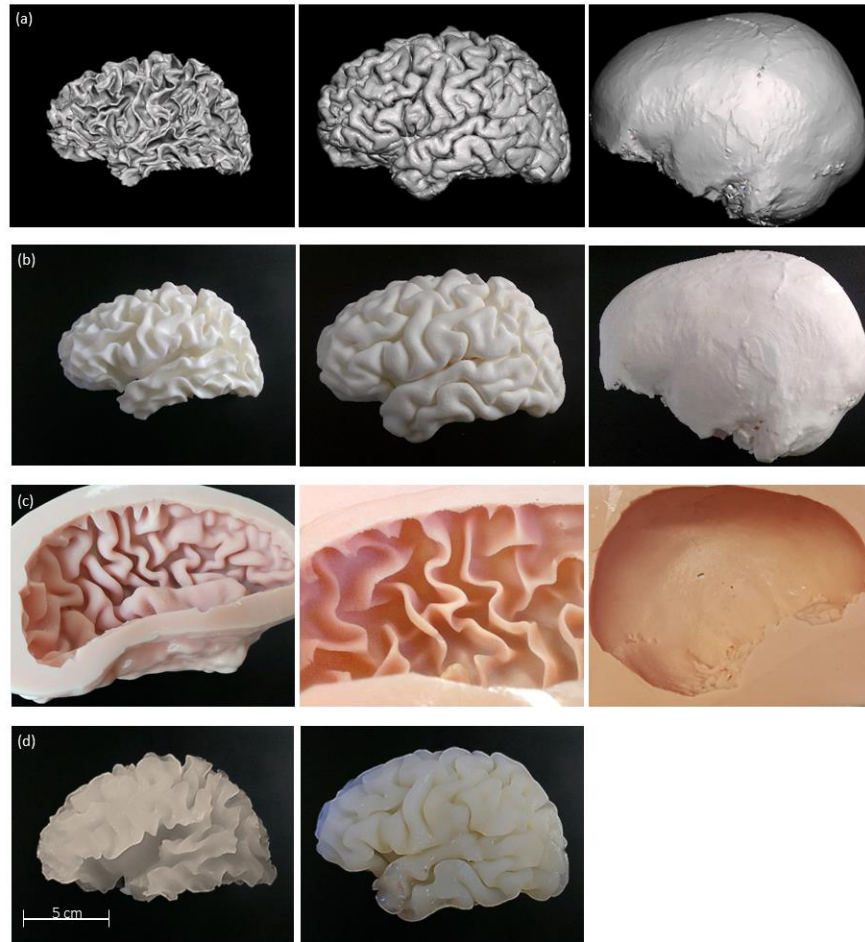


Fig. 3: Phantom shape. The surface meshes (a) were extracted from the template brain, 3D-printed with PLA (b), and brushed on with silicone to create flexible molds (c). Those molds were then used to shape the WM and GM gels (d). The left column represents the WM, the middle the GM and the last one the skull. Note that the GM gel (d, middle column) includes the WM gel (d, left column).

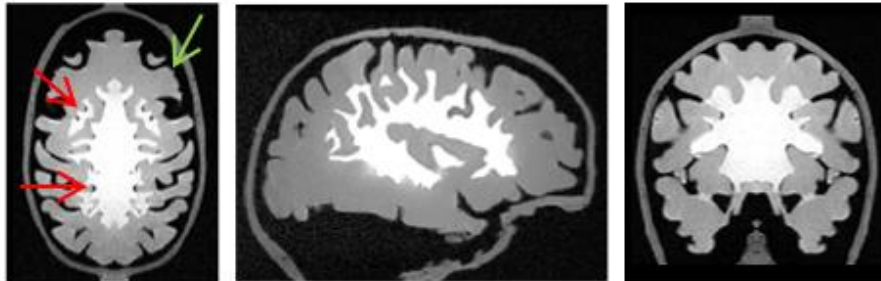


Fig. 4: MRI of the complete phantom. Axial (left), sagittal (middle), and coronal (right) view of the phantom mirrored along the midline. The 3D MPRAGE image shows the complexity of both WM and GM shapes achieved by the phantom and the homogeneity of the T1 relaxation times in the two compartments. Some cavities are observed in the WM/GM interface (red arrows) and a limited number of gyri are characterized by an unusual profile because they were not completely filled or broke off after removing the GM silicone mold (green arrow).

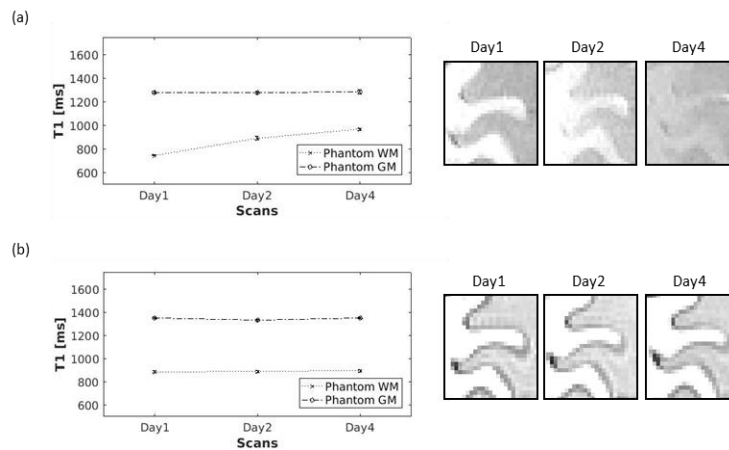


Fig. 5: Phantom interfaces and stability. T1 relaxation times and 3D MPRAGE images of Phantom2 (a) and Phantom3 (b) are shown. At Day1, both phantoms are characterized by T1 relaxation times and folding patterns comparable to those in the healthy volunteer. The wax layer applied between WM and GM in Phantom3 creates a distinguishable interface in the 3D MPRAGE image. At Days2 and 4, diffusion through the applied varnish layer in Phantom2 led to a decrease in the contrast difference between WM and GM. In Phantom3, the wax layer completely prevented the diffusion of $MnCl_2$, keeping the two compartments well separated during the observed time window of 4 days.

Discussion and Conclusions

The design and construction of the anthropomorphic physical phantoms described here were motivated by the need of providing a unique framework to test factors linked to image processing or reconstruction in conjunction with elements related to image acquisition. T1 relaxation times were reproduced using an agar gel doped with the MR contrast medium MnCl_2 and brain folding patterns were simulated using flexible silicone molds. We showed that our design does not require additional structures (i.e. chambers) to support and shape the geometry of the compartments, but at the same time, we demonstrated the need of some kind of water-repellent layer to prevent the diffusion of contrast agent. We here proposed two hydrophobic coatings that could be applied depending on the application.

The phantom features are normally dictated by the applications. Having anatomy-mimicking phantoms available for brain MRI is an important step forward for overcoming the limitations of both current physical and digital phantoms in specific applications. In this regard, a few physical phantoms with increased similarity to the real counterpart, both in terms of tissue intensities and shape, have been developed in the past years. For example, Curtis et al. (3) built 14 slices of a brain mimicking T1 and T2 tissue parameters, magnetic RF field behavior, and coil loading effects. WM and GM T1 and T2 relaxation times were reproduced using an agarose gel doped with gadolinium chloride, while each slice shape was determined from plastic compartments that were filled with the tissue-mimicking solutions. Kuss et al. (4) developed a head phantom for MRI and computed tomography (CT) to verify subdural electrode localization tools in epilepsy surgery. The phantom consisted of a skull, an adipose layer for the skin, and a deformable brain including WM and GM compartments (gelatin doped with copper sulfate for MRI, and barium sulfate for CT contrast), and CSF (agarose). Surfaces of WM and GM were imitated by manually cutting the gels to resemble the sulci structures. Another multipurpose phantom was proposed by Fujimoto et al. (5) who printed chambers from polylactic acid for WM, GM, and CSF and filled them with three different brain mimicking solutions composed of agar and manganese chloride (MnCl_2) to calibrate imaging methods. Very recently, Gopalan et al. (6) constructed a 3D-printed, anatomy-mimicking, quantitative MRI phantom for validating spatio-temporal models. Similar to the previous phantoms, brain-mimicking gels were composed of agar with nickel chloride and MnCl_2 , and 3D-printed polylactic acid chambers were filled with the corresponding solutions to mimic GM, WM and CSF. Among those existing physical phantoms, to our knowledge, these are the first prototypes built as 3D volume, characterized by such high degree of structural similarity with the real counterpart, without additional support (e.g. chambers) to maintain the shape and with a demonstrated stability of the properties over a short period of time.

Thanks to these features, the proposed phantom design could provide a unique framework to assess the impact of different steps related to image acquisition and processing on biomarkers while excluding completely the effect of biological and pathological variability. For example, the presence of WM and GM compartments and their relative complex geometries could be used to evaluate the accuracy of brain segmentation, cortical thickness estimation, shape analysis as well as the impact of different scanners and/or different imaging sequences on the accuracy of those analyses (e.g. 11). Another example of application could be in experimentally confirming the accuracy of magnetic resonance elastography (12) in detecting the *in vivo* brain tissue mechanical properties. Last but not least, our phantom could also be used in simulating longitudinal experiments. For instance, two phantoms could be built: one 'normal', and one, instead, in which physiological process occurring in pathologies, such as global or regional atrophy, lesions, and tumors, are replicated. In this case, an assessment on how, for example, different image

reconstruction techniques, compressed sensing (13), or MRI fingerprinting (14), are able to properly catch the changes simulated in the phantom could be carried out.

There are some limitations in the current phantom prototype that could be addressed in future work. First, some structural simplifications were adopted. Since only the left brain hemisphere was constructed, the phantom position within the RF coil is different from a patient scan. Then, the WM was shrunk to facilitate the assembling of the two compartments, but this step led to a thickening of the cerebral cortex. Finally, this phantom was developed focusing on T1, because this is the contrast most frequently used for morphometric measurements (for example in brain tissue segmentation, 15). However, based on this concept, further tissue properties could be added to it, for example by using a mixture of contrast agents with different T1 and T2 relaxivities to also achieve physiologically accurate T2 properties.

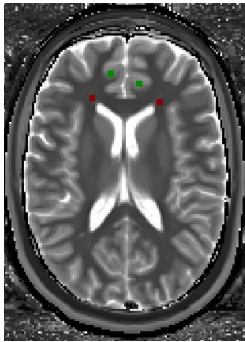
In conclusion, an innovative framework that allows the production of brain phantoms with realistic shape and homogenous T1 contrast of WM and GM has been created. The structural and intensity properties of the proposed phantoms make it useful in evaluating and validating steps from image acquisition to image processing. Moreover, the modular design makes it easily adjustable to a variety of applications.

References

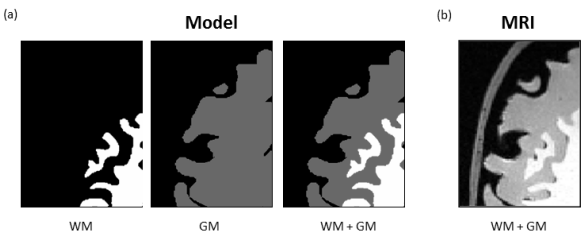
1. Kasten JA, Vetterli T, Lazeyras F, Van De Ville D. 3D-printed Shepp-Logan phantom as a real-world benchmark for MRI. *Magn. Reson. Med.* 2016;75:287–294.
2. DeWerd LA, Kissick M. The phantoms of medical and health physics : devices for research and development. 2013.
3. Curtis AT, Villemare LE, Gilbert KM, Menon RS. A head mimicking phantom for 7T, matched for tissue parameters, B1+ behavior, and coil loading effects. In: *ISMRM 18th Annual Meeting & Exhibition*. Stockholm, Sweden; 2010.
4. Kuss J, Wagner S, Meyer T, Kirsch M, Werner A, Schackert G, von Kummer R, Morgenstern U. A head phantom prototype to verify subdural electrode localization tools in epilepsy surgery. *Neuroimage* 2011;54:256–262.
5. Fujimoto K, Robertson T V., Douet V, Garmire DG, Stenger VA. A Structurally Anthropomorphic Brain Phantom. In: *ISMRM 23th Annual Meeting & Exhibition*. Toronto, Ontario, Canada; 2015.
6. Gopalan K, Tamir J, Arias A, Lustig M. Toward 3D Printed, Anatomy-Mimicking, Quantitative MRI Phantoms. In: *ISMRM 25th Annual Meeting & Exhibition*. Honolulu, Hawaii, USA; 2017.
7. Chen SJ-S, Hellier P, Marchal M, Gauvrit J-Y, Carpentier R, Morandi X, Collins DL. An anthropomorphic polyvinyl alcohol brain phantom based on Colin27 for use in multimodal imaging. *Med. Phys.* 2012;39:554–561.
8. Khan AF, Drozd JJ, Moreland RK, Ta RM, Borrie MJ, Bartha R, Alzheimer’s Disease Neuroimaging Initiative. A novel MRI-compatible brain ventricle phantom for validation of segmentation and volumetry methods. *J. Magn. Reson. Imaging* 2012;36:476–482.
9. Brown RW, Cheng Y-CN, Haacke EM, Thompson MR, Venkatesan R. *Magnetic resonance imaging : physical properties and sequence design*. Wiley; 2014.
10. Ploch CC, Mansi CSSA, Jayamohan J, Kuhl E. Using 3D Printing to Create Personalized Brain Models for Neurosurgical Training and Preoperative Planning. *World Neurosurg.* 2016;90:668–674.
11. Clark KA, Woods RP, Rottenberg DA, Toga AW, Mazziotta JC. Impact of acquisition protocols and processing streams on tissue segmentation of T1 weighted MR images. *Neuroimage* 2006;29:185–202.
12. Mariappan YK, Glaser KJ, Ehman RL. Magnetic resonance elastography: a review. *Clin. Anat.* 2010;23:497–511.
13. Geethanath S, Reddy R, Konar AS, Imam S, Sundaresan R, D R RB, Venkatesan R. Compressed sensing MRI: a review. *Crit. Rev. Biomed. Eng.* 2013;41:183–204.
14. Doerr A. Imaging: Fingerprinting with MRI. *Nat. Methods* 2013;10:380–381.
15. Despotović I, Goossens B, Philips W. MRI segmentation of the human brain: challenges, methods, and applications. *Comput. Math. Methods Med.* 2015;2015:1–23.

Supplementary Material for Review

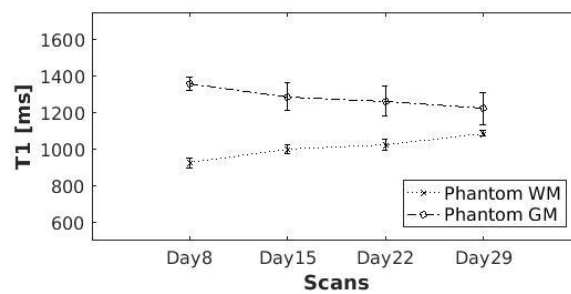
Supporting Table S1: Sample compositions for calibration curve.						
SAMPLE	1	2	3	4	5	6
Volume 2.5 mM MnCl ₂ (ml)	0	0.5	1	2	2.6	3
Volume of 0.6 % agar (ml)	50	49.5	49	48	47.4	47
Total sample volume (ml)	50	50	50	50	50	50
Concentration MnCl ₂ (mM)	0	0.03	0.05	0.1	0.13	0.15



Supporting Fig. S1: T1 mapping in template brain. T1 mapping was performed using an inversion recovery spin echo sequence (2D IR-TSE, axial slice plane, resolution = 1.3x1.3 mm², slice thickness = 5 mm, TR = 6000 ms, TE = 12 ms, TI = 30/50/100/200/500/750/1000/2000 ms, Echo train length = 7, α = 180°, total scan time = 12 min 48 s). In the T1 brain maps, average intensities within WM (2 ROIs of dimensions 3x3mm²; red squares) and GM (2 ROIs of dimensions 3x3 mm²; green squares) were computed.



Supporting Fig. S2: Geometric accuracy. (a) The created surface meshes of WM and GM were converted back to the original image format (i.e. stl-to-nifti conversion) to generate binary masks of both WM and GM. (b) These masks were then compared in a slice-by-slice fashion to the structural phantom scans.



Supporting Fig. S3: Phantom long-term stability. Diffusion of contrast agent between the two compartments starts to set in at Day8. The contrast between the two phantom compartments then diminishes constantly within the observed time window of one month.

Evaluation of Automated Brain MR Image Segmentation Methods: A Physical Phantom Study

Anna Altermatt ^{a,b}, Francesco Santini ^{a,d}, Xeni Deligianni ^{a,d}, Stefano Magon ^{b,c}, Till Sprenger ^e, Ludwig Kappos ^{a,c},
Philippe Cattin ^a, Jens Wuerfel ^{a,b}, Laura Gaetano ^{b,c}

Affiliations:

^a Department of Biomedical Engineering, University of Basel, Switzerland

^b Medical Image Analysis Center (MIAC) AG, Basel, Switzerland

^c Neurologic Clinic and Policlinic, University Hospital of Basel, Switzerland

^d Division of Radiological Physics, Department of Radiology, University Hospital of Basel, Switzerland

^e Department of Neurology, DKD HELIOS Klinik Wiesbaden, Germany

Corresponding author:

Laura Gaetano

Medical Image Analysis Center (MIAC) AG

Marktgasse 8

4051 Basel

Switzerland

Mail: laura.gaetano@miac.ch

Phone: +41 61 265 43 18

Abstract

Background: Validation of results obtained by magnetic resonance (MR) image segmentation techniques are essential to their introduction in clinical protocols. However, the validation of segmentation algorithms still remains an open issue, as the determination of the actual ground truth is not feasible *in vivo*.

Aim: Validating the quantification of white matter (WM), grey matter (GM), and total brain volumes of two commonly used automatic algorithms using a ground truth derived from a multimodal physical brain phantom.

Methods: The anthropomorphic phantom was built with two agar gels that resembled the WM and GM compartments of a real brain tissue, both in terms of signal intensity characteristics and shape. The ground truth volumes of WM, GM and total brain were derived from high-resolution computed tomography (CT) images. T1-weighted MR images acquired at 3T were used as input for automatic segmentation carried out using FMRIB Software Library (FSL v5.0; FMRIB's Automated Segmentation Tool, FAST) and Statistical Parametric Mapping (SPM v12; 'Segment'). Accuracy of the two algorithms was assessed by computing the percentage volume difference (VD) between the segmentation output and the ground truth for each compartment. Two additional T1-weighted MR images were acquired (without/with repositioning) to assess the scan-rescan reliability by comparing the computed volumes of those two scans to the ones calculated for the first scan.

Results: The ground truth volumes derived from the CT scan were: 237.60 mL for the WM, 737.92 mL for GM, and 975.52 mL for the total brain. FSL and SPM performed comparably in terms of accuracy and reliability. For the accuracy, FSL slightly prevailed in computing the volumes of WM, GM, and whole brain, yielding lower absolute values of VD than SPM ($VD_{FSL-WM} = 3.4\%$, $VD_{FSL-GM} = 2.3\%$, $VD_{FSL-Brain} = 2.6\%$; $VD_{SPM-WM} = -4.4\%$; $VD_{SPM-GM} = -2.6\%$; $VD_{SPM-Brain} = -3.0\%$). Regarding the reliability, instead, SPM provided slightly less variable results compared to FSL ($VD_{SPM} \leq 0.5\%$ without repositioning, $VD_{SPM} \leq 1.0$ with repositioning; $VD_{FSL} \leq 0.2\%$ without repositioning; $VD_{FSL} \leq 2.0$ with repositioning).

Discussion and Conclusions: The brain phantom, with its structural and intensity properties, allowed the validation of two popular automatic segmentation algorithms. Since the variability in computed volumes between the two software packages is comparable to the atrophy rates observed in aging and pathological conditions, our study underlines the huge impact of software choices for quantifying brain volumes.

Keywords: Segmentation, white matter, gray matter, FSL, SPM, physical phantom, T1

Introduction

Magnetic resonance (MR) based brain tissue segmentation is a fundamental tool for *in vivo* quantitative evaluation of brain atrophy caused by normal aging and pathological conditions (Despotović et al., 2015).

Two of the most commonly used software packages for this purpose are encoded in Statistical Parametric Mapping (SPM; <http://www.fil.ion.ucl.ac.uk/spm/>; Ashburner and Friston, 2005) and FMRIB software library (FSL; <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki>; Zhang et al., 2001). Both include automated segmentation routines that utilize iterative approaches to classify brain images into the three tissue classes white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). Besides the basic intensity-based classifications, various refinements are encoded in the standard algorithms of the two software. While the FSL algorithm borrows local spatial information from neighboring voxels (Zhang et al., 2001), the SPM algorithm relies on prior atlas information (Ashburner and Friston, 2005) to improve segmentation accuracy and robustness.

Variability in computed volumes has been shown to arise from factors related to the image acquisition or processing (Klauschen et al., 2009). Other sources of variability include physiological (circadian) variations in the hydration status and cardiovascular risk factors, which are associated with reduced brain volume (Tur et al., 2018). Most importantly, the choice of the segmentation algorithm was shown to have the biggest impact on computed WM and GM volumes (Clark et al., 2006). Since discrepancies between the different software may cause variability in computed brain volumes - which is comparable to the atrophic volume loss caused by neurodegeneration in aging or diseases - these might limit the power of tissue quantification for monitoring changes in individual patients over time (Klauschen et al., 2009).

For this reason, the validity of brain segmentation algorithms for MRI has been evaluated using manual segmentation and/or digital brain phantoms as “ground truth”. Overall, existing validation studies produced conflicting results and no clear consensus could be found on the accuracy and reliability of the underlying algorithms (Klauschen et al., 2009; Eggert et al., 2012; Kazemi and Noorizadeh, 2014; Mendrik et al., 2015; Tudorascu et al., 2016). Despite the fact that these studies could provide an indication on which segmentation algorithm is characterized by high accuracy and reliability, a definite answer is challenging to formulate due to the challenging task of defining a “ground truth”. Indeed, gold standard manual segmentations are very time-consuming with a not negligible intra- and inter-operator variability (Lampert et al., 2016). Digital brain phantoms are more flexible and easy to implement, but they might fail to incorporate all acquisition-related factors into the signal model (Kasten et al., 2016).

Hence, we proposed to validate the FSL and SPM algorithms with a novel, multimodal and anatomical physical phantom (Altermatt et al., 2017). This phantom simulates the WM and GM compartments of the human brain in terms of T1 relaxation times, radiopacity, and shape. Being a physical object, it can be scanned using the MRI technology, therefore incorporating the real acquisition process in the evaluation, as well as imaged with high resolution computed tomography (CT), thus providing an accurate and objective “ground truth”.

Materials and Methods

2.1. Physical Brain Phantom

The single-hemisphere brain phantom was built according to the procedure described in (Altermatt et al., 2017). Briefly, the phantom mimics the brain WM and GM compartments, both in terms of T1 relaxation times and shape. The former was achieved with two 0.6% agar gels (Morga AG, Ebnat-Kappel, Switzerland) doped with, 0.12 and 24 mg/mL of MR contrast agent respectively (Manganese chloride, MnCl_2 ; Sigma-Aldrich Chemie GmbH, Steinheim, Germany). The latter was obtained using a dedicated molding-casting procedure that allowed the replication of the complex 3D geometry of the two compartments. For this study, 24 mg/mL CT contrast agent (Iopromidium; Bayer AG, Leverkusen, Germany) was also added to the WM agar solution to obtain distinguishable attenuation coefficients in the CT scan.

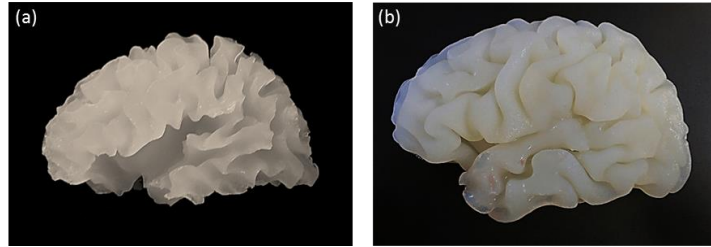


Figure 1: Anthropomorphic Phantom. (a) Phantom WM compartment. (b) Phantom GM compartment (enclosing the WM).

2.2. Ground truth volumes

For the computation of the ground truth volumes, the described phantom design allowed us to take advantage of the high spatial and temporal resolution of the CT technology. The phantom was thus first scanned using CT (Somatom Edge, Siemens Healthineers, Erlangen, Germany resolution = $0.44 \times 0.44 \times 0.25 \text{ mm}^3$, eff. mAs of 270, 0.6 mm slice thickness, J30s/3 reconstruction kernel, 226 mm FOV). After mirroring the resulting image at the mid-sagittal plane, the number of voxels belonging to the WM and GM compartments was computed applying a threshold at 1170 HU: if the voxel intensity was below this value, then the voxel was considered as part of GM tissue, if equal or above, as WM. The number of voxels of the two compartments was then multiplied by the voxel dimension to obtain the ground truth WM and GM volumes. The ground truth whole brain volume was computed as the sum of WM and GM volumes.

2.3. MR image acquisition

The MRI protocol included a T1-weighted imaging sequence (3D MPRAGE, sagittal slice plane, resolution = $0.82 \times 0.82 \times 1.00 \text{ mm}^3$, TR = 1570 ms, TI = 900 ms, TE = 1.34 ms, $\alpha = 8^\circ$, scan time = 6 min 38 s) acquired with a 20-channel head coil at a 3 Tesla clinical scanner (Magnetom Prisma, Siemens Healthineers, Erlangen, Germany). 2D distortion correction of the images was performed according to the manufacturer's specifications by enabling the relevant parameter in the sequence options. The protocol was repeated three times within the same session: a baseline scan, a rescan without repositioning, and a rescan with repositioning of the phantom between the scans.

2.4. Image segmentation

Following the acquisition, all MRI data were first mirrored along the mid-sagittal plane of the phantom. The images were then used as input for FSL and SPM to compute WM, GM, and whole brain volumes. The FSL segmentation algorithm was run using the default parameters. A threshold of 0.5 was applied to the resulting probability masks to generate binary masks for each class. Volumes of WM and GM were computed by multiplying the number of voxels in each mask by the voxel dimensions. Unlike FSL, SPM uses tissue probability maps as priors for the segmentation (Ashburner and Friston, 2005). To account for the anatomical features of the phantom (Altermatt et al., 2017), only three priors were selected (the WM, GM, and the background one) and modified in the following way: the ventricles in the WM prior were filled; in the GM prior, the cortical thickness was increased by 20 mm and the WM prior was scaled down by the same amount. A threshold of 0.5 was applied to the probability maps derived from the segmentation, generating binary masks from which the WM and GM volumes were calculated. The whole brain volume was computed by adding together the WM and GM volumes. The same procedure was then repeated to assess the volumes in the scan-rescan experiments (without and with repositioning).

2.5. Evaluation of segmentation

The results obtained from the segmentation of the MRI were qualitatively and quantitatively compared to the ground truth derived from the CT image. A visual assessment was performed identifying misclassified voxels in the binary maps, if any. A quantitative evaluation of the results was also carried out calculating two indexes: the percentage volume difference (VD) and the Dice similarity coefficient (DSC, Dice, 1945). The VD quantified how much the volumes derived from FSL or SPM differed from the ground truth ones. It was defined as

$$VD = \frac{V_{MRI} - V_{CT}}{V_{CT}} \cdot 100,$$

where V_{MRI} is the volume obtained from the segmentation algorithms (i.e. FSL or SPM) and V_{CT} is the ground truth volume extracted from the CT scan.

Then, CT images were registered to the MRI and the transformation matrix was applied to the CT-derived whole brain, WM, and GM binary mask. To assess the spatial overlap of the binary masks, the DSC was calculated as:

$$DSC = \frac{2|A \cap G|}{|A| + |G|} \cdot 100,$$

where A is the masks derived from the automatic segmentations (i.e. by FSL or SPM) and G is the corresponding ground truth mask.

To evaluate the scan-rescan reliability, the volumes of the first scan were compared to the second and third scans (without and with repositioning, respectively) by calculating VDs in which the volumes from the first scan were used as reference.

All the analyses were done using Matlab (version R2017b; <https://www.mathworks.com/>).

Results

Acquired CT and T1-weighted MR images are shown in Figure 2. The ground truth volumes derived from the CT scan were: 237.60 mL for the WM, 737.92 mL for GM, and 975.52 mL for the whole brain. These ground truth binary masks for WM, GM, and total brain and the binary masks resulting from automatic segmentation are shown in Figure 3.

While the FSL algorithm overestimated volumes of WM, GM, and total brain by up to 3.4%, the same volumes were underestimated by up to -4.4% using SPM (Table 1). In terms of absolute values, FSL provided volumes closer to the ground truth than SPM. The visual assessment showed some false positive WM voxels in the FSL segmentation at locations within the GM area (Figure 3), which are correctly classified by the SPM algorithm. The visual assessment further showed that the FSL algorithm seems to accurately contour the outer GM border, while the same area was clearly underestimated by the SPM algorithm (Figure 3). Differences in computed volumes between the two automatic segmentation software reached discrepancies of 7.8% for the WM, 4.9% for the GM, and 5.6% for the total brain.

Both software packages were characterized by a very good and comparable scan-rescan reliability, as shown by the scan-rescan assay (without/with repositioning) (Table 2). In the scan-rescan experiment with repositioning, higher differences were generally observed (Table 2) reaching a maximal value of 2% for the WM volume calculated by FSL.

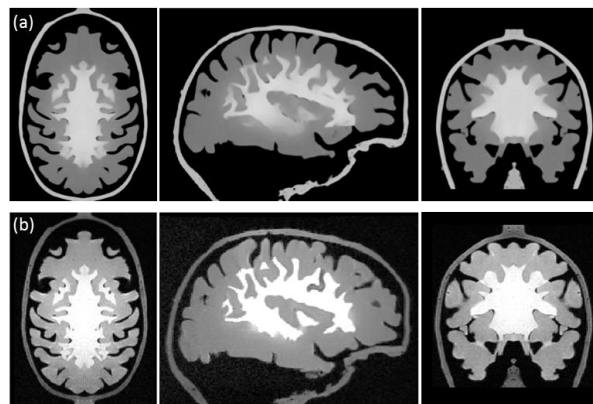


Figure 2: Multimodal phantom images. (a) Acquired high-resolution CT images used to derive phantom ground truth volumes. (b) Acquired T1-weighted MRI used for automatic segmentation by the software FSL and SPM. Axial (left column), sagittal (middle column), and coronal (right column) slice planes.

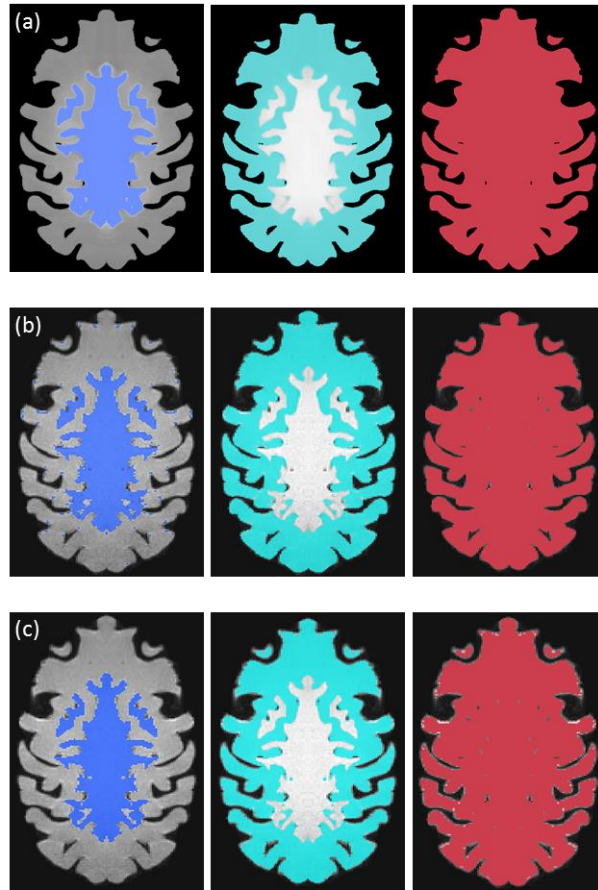


Figure 3: Evaluation of segmentation. Figure depicting the binary masks of WM (left column), GM (middle column), and total brain (right column) derived from (a) ground truth derived from CT, (b) FSL FAST, and (c) SPM 'Segment' segmentations.

Table 1: Accuracy of segmentation

	Total Brain		WM		GM	
	<i>VD [%]</i>	<i>DSC [%]</i>	<i>VD [%]</i>	<i>DSC [%]</i>	<i>VD [%]</i>	<i>DSC [%]</i>
FSL	2.55	96	3.37	92	2.29	94
SPM	-3.03	93	-4.39	94	-2.59	94

VD: Volumetric difference; DSC: Dice similarity coefficient

Table 2: Scan-rescan reliability

	Total Brain	WM	GM
	<i>VD [%]</i>	<i>VD [%]</i>	<i>VD [%]</i>
FSL			
Rescan (w/o repositioning)	-0.12	+0.08	-0.19
Rescan (with repositioning)	+0.61	+2.04	+0.14
SPM			
Rescan (w/o repositioning)	+0.01	-0.49	+0.17
Rescan (with repositioning)	+0.38	+0.96	+0.20

VD: Volumetric difference

Discussion

We validated the quantification of whole brain, WM, and GM volumes by two commonly used automatic algorithms, FSL and SPM, using an objective and accurate ground truth derived from a multimodal physical brain phantom.

Regarding the accuracy, in the current study, FSL slightly prevailed in computing the volumes of WM, GM, and whole brain. This is in agreement with Klauschen et al. (2009) and Eggert et al. (2012), but in discordance with Tsang et al. (2005), Kazemi et al. (2014), Mendrik et al. (2015), and Heinen et al. (2016). The reasons for those differences could be various, ranging from which type of ground truth was used (i.e. digital simulations, real images with manual segmentation, or both), software version, field strength and type of sequences (Clark et al., 2006; Heinen et al., 2016). For example, both Tsang et al. (2005) and Kazemi et al. (2014) used previous versions of the segmentation algorithms, the ground truth was digital simulated brains and real images with manual tracing acquired at 1.5 Tesla. Mendrik et al. (2015) adopted the same versions of FSL and SPM as in the current study, the real images were also acquired at 3 Tesla, but the segmentation was performed on 2D T1 as input for SPM and FSL. Similarly, Heinen et al. (2016) downsampled the 3D T1 images to make the manual segmentation of WM and GM feasible. Finally, Tudorascu et al. (2016) preferred SPM based exclusively on a quality assessment by two neurologists performed on ten specific GM regions of interest.

Regarding the reliability, overall, SPM provided more consistent results compared to FSL. Also Klauschen et al. (2009) and Eggert et al. (2012) showed that FSL had more variable results, suggesting that FSL is more sensitive to noise factors than SPM. However, we obtained a lower GM volume difference between the two scans than Eggert et al. (2012) with both segmentation algorithms. Tudorascu et al. (2016), instead, showed a higher agreement in the segmentation of GM by SPM and of WM by FSL, that could be in agreement with our results that showed a lower GM volume difference by SPM and WM volume difference by FSL, instead. Moreover, our volume difference in the whole brain segmented by FSL is also included in the error range ($0.4\pm6\%$) reported by (Smith et al., 2002) in a scan-rescan experiment. However, despite being in this ranges, our values are comparable to, in the case of SPM, or exceeding, in the case of FSL, the annual rate of brain volume difference reported in neurodegenerative disorders, such as multiple sclerosis (De Stefano et al., 2010). This might especially limit the power of these methods for monitoring changes in individual patients over time in longitudinal studies (Klauschen et al., 2009). Indeed, these findings underline the tremendous effect of technical confounders in the analyses of brain volumes and that these considerations should be taken into account when interpreting existing studies, especially if in small cohorts.

Local discrepancies in the segmentation accuracy between the two software packages were detected. The FSL segmentation has shown many false positive WM voxels within the GM region. These voxels were most likely misclassified by the FSL algorithm because of their higher T1-weighted intensities compared to the average GM intensity. Since FSL primarily relies on the intensity distribution of the images and the local spatial information surrounding a voxel, it might be, in fact, more sensitive to imaging artefacts and noise (Zhang et al., 2001). The SPM algorithm, instead, may be less influenced by that because of its adoption of prior atlas information, which might reduce classification errors by assigning a low probability to the WM at atlas-defined locations of the GM class (Ashburner and Friston, 2005). Despite reducing the number of misclassifications in this way, the prior atlas information encoded in the SPM algorithm does not seem to strongly improve the overall segmentation accuracy.

Unlike FSL, the SPM algorithm was shown to slightly underestimate all volumes, which could mainly be linked to the misclassification of voxels affected by partial volume effects.

Compared to the “gold standard” established using manual segmentation, the automatic use of a physical phantom that we proposed here is more time-efficient, more accurate and not subjected to intra- inter-operator bias (Despotovic et al., 2015). Indeed, in the high-resolution phantom CT images, the ground truth can be established upon simple thresholding. This segmentation step is fully automatic and therefore easily reproducible. Compared to digital phantoms, this physical phantom offers the advantage of allowing to assess the true effect of choices related to image acquisition (i.e. type of scanner/RF coil, sequence specifications, artefacts, patient positioning), which can only be simulated for artificial images (Despotovic et al., 2015). For example, methods that are optimized based on numerical simulations may fail or introduce systematic bias when presented with real scanner data if specific factors are discounted or overlooked. Although analytical Fourier expressions may be bound for a variety of objects, introducing additional experimental factors – such as for example arbitrary field perturbations - may lead to mathematically intractable situations in digital simulations (Kasten et al., 2016). Among existing physical phantoms mimicking structure and T1 relaxation times of WM and GM (Curtis et al., 2010; Kuss et al., 2011; Khan et al., 2012; Fujimoto et al., 2015; Gopalan et al., 2017), to our knowledge, this is the first prototype characterized by such high degree of structural similarity with real brain anatomy and does not require additional chambers to maintain the shape.

However, despite this phantom being a step further in the definition of a ground truth, it is still a simplification of a real brain. The lack of a surrogate for the cerebrospinal fluid causes susceptibility artefacts that translated into a hyperintense margin at the outer border of the GM. Moreover, the very pronounced cortical thickness was fundamental for building the phantom (Altermatt et al., 2017), but it reduces the impact of partial volume effects on the GM compartment. Furthermore, the homogeneity of the gels used for the phantom in terms of T1 relaxation times influences the intensity histogram, simplifying to some extent the segmentation by both algorithms.

Future studies using the same anthropomorphic phantom could analyze the impact of different choices related to the image acquisition on computed total or regional brain volumes. For example, volume differences resulting from scanner changes (across field strengths and vendors), from modifications to the imaging sequences, or from phantom positioning could be quantified.

In conclusion, the multimodal physical brain phantom provided an accurate and reliable ground truth for the validation of segmentation algorithms. While FSL prevailed in terms of accuracy, SPM showed a better performance in terms of reliability. Since volume differences between scans and software were shown to reach discrepancies comparable to atrophy, this study underlines the tremendous impact of technical confounders in the quantification of total and regional brain volumes.

References

- Altermatt, A., Santini, F., Deligianni, X., Magon, S., Cattin, P., Wuerfel, J., Gaetano, L., 2017. Novel design of a 3D-printed anthropomorphic brain phantom for segmentation validation in magnetic resonance imaging, in: ISMRM 25th Annual Meeting & Exhibition. Honolulu, Hawaii, USA.
- Ashburner, J., Friston, K.J., 2005. Unified segmentation. *Neuroimage* 26, 839–851.
- Clark, K.A., Woods, R.P., Rottenberg, D.A., Toga, A.W., Mazziotta, J.C., 2006. Impact of acquisition protocols and processing streams on tissue segmentation of T1 weighted MR images. *Neuroimage* 29, 185–202.
- Curtis, A.T., Villemaire, L.E., Gilbert, K.M., Menon, R.S., 2010. A head mimicking phantom for 7T, matched for tissue parameters, B1+ behavior, and coil loading effects, in: ISMRM 18th Annual Meeting & Exhibition. Stockholm, Sweden.
- De Stefano, N., Giorgio, A., Battaglini, M., Rovaris, M., Sormani, M.P., Barkhof, F., Korteweg, T., Enzinger, C., Fazekas, F., Calabrese, M., Dinacci, D., Tedeschi, G., Gass, A., Montalban, X., Rovira, A., Thompson, A., Comi, G., Miller, D.H., Filippi, M., 2010. Assessing brain atrophy rates in a large population of untreated multiple sclerosis subtypes. *Neurology* 74, 1868–1876.
- Despotović, I., Goossens, B., Philips, W., 2015. MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. *Comput. Math. Methods Med.* 2015, 1–23.
- Dice, L.R., 1945. Measures of the Amount of Ecologic Association Between Species. *Ecology* 26, 297–302.
- Eggert, L.D., Sommer, J., Jansen, A., Kircher, T., Konrad, C., 2012. Accuracy and reliability of automated gray matter segmentation pathways on real and simulated structural magnetic resonance images of the human brain. *PLoS One* 7, e45081.
- Fujimoto, K., Robertson, T. V., Douet, V., Garmire, D.G., Stenger, V.A., 2015. A structurally anthropomorphic brain phantom, in: ISMRM 23th Annual Meeting & Exhibition. Toronto, Ontario, Canada.
- Gopalan, K., Tamir, J., Arias, A., Lustig, M., 2017. Toward 3D printed, anatomy-mimicking, quantitative MRI phantoms, in: ISMRM 25th Annual Meeting & Exhibition. Honolulu, Hawaii, USA.
- Kasten, J.A., Vetterli, T., Lazeyras, F., Van De Ville, D., 2016. 3D-printed shepp-logan phantom as a real-world benchmark for MRI. *Magn. Reson. Med.* 75, 287–294.
- Kazemi, K., Noorizadeh, N., 2014. Quantitative Comparison of SPM, FSL, and Brainsuite for Brain MR Image Segmentation. *J. Biomed. Phys. Eng.* 4, 13–26.
- Khan, A.F., Drozd, J.J., Moreland, R.K., Ta, R.M., Borrie, M.J., Bartha, R., Alzheimer’s Disease Neuroimaging Initiative, 2012. A novel MRI-compatible brain ventricle phantom for validation of segmentation and volumetry methods. *J. Magn. Reson. Imaging* 36, 476–482.
- Klauschen, F., Goldman, A., Barra, V., Meyer-Lindenberg, A., Lundervold, A., 2009. Evaluation of automated brain MR image segmentation and volumetry methods. *Hum. Brain Mapp.* 30, 1310–1027.
- Kuss, J., Wagner, S., Meyer, T., Kirsch, M., Werner, A., Schackert, G., von Kummer, R., Morgenstern, U., 2011. A head phantom prototype to verify subdural electrode localization tools in epilepsy surgery. *Neuroimage* 54, 256–262.
- Lampert, T.A., Stumpf, A., Gancarski, P., 2016. An empirical study into annotator agreement, ground truth estimation, and algorithm evaluation. *IEEE Trans. image Process.* 25, 2557–2572.

- Mendrik, A.M., Vincken, K.L., Kuijf, H.J., Breeuwer, M., Bouvy, W.H., de Bresser, J., Alansary, A., de Bruijne, M., Carass, A., El-Baz, A., Jog, A., Katyal, R., Khan, A.R., van der Lijn, F., Mahmood, Q., Mukherjee, R., van Opbroek, A., Paneri, S., Pereira, S., Persson, M., Rajchl, M., Sarikaya, D., Smedby, Ö., Silva, C.A., Vrooman, H.A., Vyas, S., Wang, C., Zhao, L., Biessels, G.J., Viergever, M.A., 2015. MRBrainS challenge: online evaluation framework for brain image segmentation in 3T MRI scans. *Comput. Intell. Neurosci.* 2015, 1–16.
- Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., De Stefano, N., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage* 17, 479–89.
- Tudorascu, D.L., Karim, H.T., Maronge, J.M., Alhilali, L., Fakhran, S., Aizenstein, H.J., Muschelli, J., Crainiceanu, C.M., 2016. Reproducibility and bias in healthy brain segmentation: comparison of two popular neuroimaging platforms. *Front. Neurosci.* 10, 503.
- Tur, C., Moccia, M., Barkhof, F., Chataway, J., Sastre-Garriga, J., Thompson, A.J., Ciccarelli, O., 2018. Assessing treatment outcomes in multiple sclerosis trials and in the clinical setting. *Nat. Rev. Neurol.* 14, 75–93.
- 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.

7 Discussion and Conclusions

Despite the central role of MRI in assessing and quantifying brain lesions and atrophy, clinico-radiological associations are often weak, limiting the use of MRI-derived markers in the clinical routine or as endpoints in clinical trials [5–7]. The work conducted as part of this thesis was motivated by the need of bridging the gap between clinical and radiological disease manifestations. The two approaches presented here, addressed two different topics, which might underlie this mismatch: (A) In a voxel-based lesion-symptom mapping study, we aimed at strengthening *lesion-symptom* associations by the identification of strategic lesion locations; (B) With the construction of an anatomical physical phantom we aimed at providing a "ground truth" for the validation of atrophy measurements, which will enable addressing the robustness and reliability of *atrophy-symptom* associations.

Lesion-symptom association: Voxel-based lesion-symptom mapping was performed in two large cohorts: a dataset of 2348 relapsing-remitting MS patients, and a population-based cohort of 1017 elderly subjects. For both studies, an automatized registration pipeline for alignment of the T2-weighted lesion masks to standard stereotaxic space was implemented. In a second step, a voxel-based statistical approach to relate lesion location to different clinical measures was encoded.

In the MS lesion mapping study, significant associations between lesion location and several clinical scores were found in periventricular areas. Such lesion clusters appear to be associated with impairment of different physical and cognitive abilities, probably because they affect commissural and long projection fibres, which are relevant WM pathways supporting many different brain functions. In the SVD lesion mapping study, age, gender, several cognitive scores, and WM lesion load were shown to be significantly associated with vascular risk factors. Moreover, periventricular WM lesions affecting the same WM fibres and the caudate nucleus were identified to significantly relate to the subjects' cerebrovascular risk profiles, while no other WM locations were found to be associated with cognitive impairment in this population of elderly, but on average cognitively preserved adults.

As becomes obvious from the obtained lesion distributions in these two large cohorts, in both diseases, WM lesions occur more frequently around the ventricles than elsewhere. For this reason, with conventional T2-weighted MRI it might be difficult to distinguish between the two disease pathologies, especially since they often coexist. In SVD, sparing of the U-fibers, relative sparing of the corpus callosum, temporal lobe and cerebellum, more central involvement of brain stem, presence of lacunar infarcts and microbleeds and sparing of spinal cord are common findings to distinguish it from MS with conventional

MRI [36, 69]. The predilection for periventricular lesion locations in both diseases might be suggestive for a common pathophysiological mechanism underlying the two diseases. In MS, the underlying substrate for this topographic predilection is not fully understood. Some studies suggest an increased vulnerability of periventricular sites due to the presence of soluble mediators within the CSF which trigger autoimmunity [70]. Other observations suggest that the relationship between MS lesions and the cerebral vessels is important. Periventricular lesions form around subependymal veins and perivascular (venous) inflammatory infiltrates are seen in these MS lesions [70]. Moreover, focal WM lesions seem to occur in watershed areas of low arterial blood flow. Following this observation, tissue hypoxia has been suggested as common pathogenic mechanism potentially underlying both disease. For haemodynamic reasons periventricular WM may be more likely than other parts of the brain to become hypoxic, explaining the predilection for lesion formation in this region [71].

There are several reasons for the lack or non-specificity of the identified lesion-symptom associations that apply to both studies and which could be addressed in future work. First, the type of lesions considered for these analyses reflect a variety of pathophysiological processes and severity of tissue damage seen in MS and SVD. Second, other central nervous system compartments (e.g. cortical lesions) which might play a role in the development of disability, were not taken into account in these lesion-mapping approaches [5–7]. Finally, future studies could aim at implementing a lesion-mapping algorithm which allows incorporating information from neighboring voxels or which partly captures the complex network underlying brain function [14, 72]. Indeed, studies exploring the relation between lesions and clinical disability in MS have shown that the moderate clinico-radiological association is rather due to the intrinsic variability in the clinical expression of lesions at different anatomical locations than due to deficiencies in existing MRI markers or clinical rating scales [73].

In summary, significant associations between lesion location and clinical impairment in MS and cerebrovascular risk factors relevant for SVD were found in periventricular areas, reflecting damage to commissural and large projection fibres. No other lesion locations have shown a significant association to the cognitive scores considered in the SVD lesion mapping. We conclude that voxel-based lesion-symptom mapping of T2-weighted WM lesions may in some cases reveal a stronger lesion-symptom association to clinical impairment compared to conventional lesion assessments. Future lesion-mapping studies should try to partly capture the complex network underlying brain function by identifying associations between lesional patterns and clinical impairment.

Atrophy-symptom association: With the construction of an anatomical physical phantom we established a "ground truth" for segmentation validation, which will enable addressing the robustness and reliability of *atrophy-symptom* associations. The constructed phantom prototype is composed of agar gels doped with MR and computed tomography contrast agents, which realistically mimic T1 relaxation times at 3 Tesla and showing distinguishable attenuation coefficients using computed tomography. Moreover, due to the design of anatomical and flexible molds, both WM and GM are character-

ized by shapes comparable to the human counterpart. Due to the increased anatomical similarity and the absence of thick physical (e.g. 3D-printed) interfaces between the compartments, the phantom allows to more realistically mimic image segmentation compared to existing phantoms.

In a proof-of-principle study, the designed phantom was used to validate automatic brain tissue segmentation by two commonly-used software tools. "Ground truth" volumes of the two phantom compartments were derived from a high-resolution computed tomography scan. Automatic image segmentation using the standard tools available in the two software FSL and SPM was done on T1-weighted phantom images. Both automatic segmentation algorithms were shown to perform comparably, while FSL was shown to be slightly more accurate in computing the volumes of WM, GM, and total brain. While the FSL algorithm slightly overestimated volumes of WM, GM, and total brain, the same volumes were slightly underestimated by SPM. Regarding the reliability of computed volumes by the two software, the two algorithms were again shown to perform comparably, while SPM provided slightly more consistent results compared to FSL. Results across software were shown to be highly variable reaching volume differences of up to 8%. These discrepancies are far beyond the yearly whole-brain or regional atrophy rates observed in MS or SVD and, in accordance with results of several previous studies [21, 22, 74, 75], our study underlines the huge impact of software choices in the analyses of brain volumes.

There are some limitations in the current phantom prototype that need to be addressed in future work. Most importantly, the cortical thickness is increased in the employed phantom, which might for example reduce the impact of partial voluming for segmentation applications. Then, since the phantom does not realistically mimic the difference in susceptibilities between GM and CSF, a hyperintense margin is visible in the T1-weighted MRI, which might slightly influence GM segmentation. Then, the current phantom prototype produces constant T1 relaxation times only during several days and future work is needed to create a phantom, which is stable within a longer time window. Finally, since we have established a modular framework for phantom construction, different brain regions could be incorporated and different physical properties (e.g. T2, PD) could be mimicked additionally.

Future envisioned applications of the constructed phantom are to study the impact of different choices related to the image acquisition on computed total or regional brain volumes. For example volume difference resulting from scanner changes (across field strengths and vendors), from modifications to the imaging sequences, or from phantom positioning could be quantified using the designed anatomical phantom.

To sum up, an anatomical physical brain phantom prototype mimicking shape and T1 of WM and GM was successfully constructed. The phantom was used in a proof-of-principle study to validate and compare two commonly-used automatic segmentation tools. To sum up, while FSL slightly overestimated volumes of WM, GM, and total brain, the same volumes were slightly underestimated by SPM. In general, results from the same software yielded reliable results across scans, while results across software were shown to be highly variable reaching volume differences beyond the yearly regional and total brain

atrophy rates. In this regard, our findings underline the central role of software choices to reliably and robustly associate MRI-derived atrophy measurements and symptoms in neurological diseases.

Bibliography

- [1] Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., Mowry, E. M., Sorensen, P. S., Tintoré, M., Traboulsee, A. L., Trojano, M., Uitdehaag, B. M., Vukusic, S., Waubant, E., Weinshenker, B. G., Reingold, S. C., and Cohen, J. A. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*, 17(2):162–173 (2018).
- [2] Pantoni, L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *The Lancet Neurology*, 9(7):689–701 (2010).
- [3] Filippi, M. Magnetic Resonance Techniques in Multiple Sclerosis. *Archives of Neurology*, 68(12):1514 (2011).
- [4] Wardlaw, J. M., Smith, E. E., Biessels, G. J., Cordonnier, C., Fazekas, F., Frayne, R., Lindley, R. I., O’Brien, J. T., Barkhof, F., Benavente, O. R., Black, S. E., Brayne, C., Breteler, M., Chabriat, H., DeCarli, C., de Leeuw, F. E., Doubal, F., Duering, M., Fox, N. C., Greenberg, S., Hachinski, V., Kilimann, I., Mok, V., van Oostenbrugge, R., Pantoni, L., Speck, O., Stephan, B. C., Teipel, S., Viswanathan, A., Werring, D., Chen, C., Smith, C., van Buchem, M., Norrving, B., Gorelick, P. B., and Dichgans, M. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *The Lancet Neurology*, 12(8):822–838 (2013).
- [5] Chard, D. and Trip, S. A. Resolving the clinico-radiological paradox in multiple sclerosis. *F1000Research*, 6:1828 (2017).
- [6] Nitkunan, A., Lanfranconi, S., Charlton, R. A., Barrick, T. R., and Markus, H. S. Brain atrophy and cerebral small vessel disease a prospective follow-up study. *Stroke*, 42(1):133–138 (2011).
- [7] Biesbroek, J. M., Weaver, N. A., and Biessels, G. J. Lesion location and cognitive impact of cerebral small vessel disease. *Clinical Science*, 131(8):715–728 (2017).
- [8] Charil, A., Zijdenbos, A. P., Taylor, J., Boelman, C., Worsley, K. J., Evans, A. C., and Dagher, A. Statistical mapping analysis of lesion location and neurological disability in multiple sclerosis: application to 452 patient data sets. *NeuroImage*, 19(3):532–544 (2003).

- [9] Sepulcre, J., Masdeu, J. C., Sastre-Garriga, J., Goñi, J., Vélez-de Mendizábal, N., Duque, B., Pastor, M. A., Bejarano, B., and Villoslada, P. Mapping the brain pathways of declarative verbal memory: Evidence from white matter lesions in the living human brain. *NeuroImage*, 42(3):1237–1243 (2008).
- [10] Sepulcre, J., Masdeu, J. C., Pastor, M. A., Goñi, J., Barbosa, C., Bejarano, B., and Villoslada, P. Brain pathways of verbal working memory: A lesion-function correlation study. *NeuroImage*, 47(2):773–778 (2009).
- [11] Vellinga, M. M., Geurts, J. J. G., Rostrup, E., Uitdehaag, B. M. J., Polman, C. H., Barkhof, F., and Vrenken, H. Clinical correlations of brain lesion distribution in multiple sclerosis. *Journal of magnetic resonance imaging*, 29(4):768–73 (2009).
- [12] Kincses, Z., Ropele, S., Jenkinson, M., Khalil, M., Petrovic, K., Loitfelder, M., Langkammer, C., Aspeck, E., Wallner-Blazek, M., Fuchs, S., Jehna, M., Schmidt, R., Vecsei, L., Fazekas, F., and Enzinger, C. Lesion probability mapping to explain clinical deficits and cognitive performance in multiple sclerosis. *Multiple Sclerosis Journal*, 17(6):681–689 (2011).
- [13] Rossi, F., Giorgio, A., Battaglini, M., Stromillo, M. L., Portaccio, E., Goretti, B., Federico, A., Hakiki, B., Amato, M. P., and De Stefano, N. Relevance of brain lesion location to cognition in relapsing multiple sclerosis. *PLoS one*, 7(11):e44826 (2012).
- [14] Ge, T., Müller-Lenke, N., Bendfeldt, K., Nichols, T. E., and Johnson, T. D. Analysis of multiple sclerosis lesions via spatially varying coefficients. *The Annals of Applied Statistics*, 8(2):1095–1118 (2014).
- [15] Duering, M., Zieren, N., Hervé, D., Jouvent, E., Reyes, S., Peters, N., Pachai, C., Opherk, C., Chabriat, H., and Dichgans, M. Strategic role of frontal white matter tracts in vascular cognitive impairment: A voxel-based lesion-symptom mapping study in CADASIL. *Brain*, 134(8):2366–2375 (2011).
- [16] Biesbroek, J. M., Kuijf, H. J., van der Graaf, Y., Vincken, K. L., Postma, A., Mali, W. P. T. M., Biessels, G. J., and Geerlings, M. I. Association between subcortical vascular lesion location and cognition: a voxel-based and tract-based lesion-symptom mapping study. The SMART-MR study. *PLoS one*, 8(4) (2013).
- [17] Duering, M., Gesierich, B., Seiler, S., and Gonik, M. Strategic white matter tracts for processing speed deficits in age-related small vessel disease. *American Academy of Neurology*, 82(22):1946–1950 (2014).
- [18] Smith, E. E., O'Donnell, M., Dagenais, G., Lear, S. A., Wielgosz, A., Sharma, M., Poirier, P., Stotts, G., Black, S. E., Strother, S., Noseworthy, M. D., Benavente, O., Modi, J., Goyal, M., Batool, S., Sanchez, K., Hill, V., McCreary, C. R., Frayne, R., Islam, S., DeJesus, J., Rangarajan, S., Teo, K., and Yusuf, S. Early cerebral small vessel disease and brain volume, cognition, and gait. *Annals of Neurology*, 77(2):251–261 (2015).

- [19] Biesbroek, J. M., van Zandvoort, M. J., Kappelle, L. J., Velthuis, B. K., Biessels, G. J., and Postma, A. Shared and distinct anatomical correlates of semantic and phonemic fluency revealed by lesion-symptom mapping in patients with ischemic stroke. *Brain Structure and Function*, 221(4):2123–2134 (2016).
- [20] Klein, A., Andersson, J. L. R., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-c., Christensen, G., Collins, L., Hellier, P., Song, P., Hyun, J., Lepage, C., Pennec, X., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R. P., Mann, J. J., and Parsey, R. V. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802 (2009).
- [21] Clark, K. A., Woods, R. P., Rottenberg, D. A., Toga, A. W., and Mazziotta, J. C. Impact of acquisition protocols and processing streams on tissue segmentation of T1 weighted MR images. *NeuroImage*, 29(1):185–202 (2006).
- [22] Klauschen, F., Goldman, A., Barra, V., Meyer-Lindenberg, A., and Lundervold, A. Evaluation of automated brain MR image segmentation and volumetry methods. *Human brain mapping*, 30(4):1310–27 (2009).
- [23] Curtis, A. T., Villemaire, L. E., Gilbert, K. M., and Menon, R. S. A Head Mimicking Phantom for 7T, Matched for Tissue Parameters, B1+ Behavior, and Coil Loading Effects. In *ISMRM 18th Annual Meeting & Exhibition*. Stockholm, Sweden (2010).
- [24] Kuss, J., Wagner, S., Meyer, T., Kirsch, M., Werner, A., Schackert, G., von Kummer, R., and Morgenstern, U. A head phantom prototype to verify subdural electrode localization tools in epilepsy surgery. *NeuroImage*, 54:256–262 (2011).
- [25] Chen, S. J.-S., Hellier, P., Marchal, M., Gauvrit, J.-Y., Carpentier, R., Morandi, X., and Collins, D. L. An anthropomorphic polyvinyl alcohol brain phantom based on Colin27 for use in multimodal imaging. *Medical Physics*, 39(1):554–561 (2012).
- [26] Khan, A. F., Drozd, J. J., Moreland, R. K., Ta, R. M., Borrie, M. J., Barth, R., and Alzheimer’s Disease Neuroimaging Initiative. A novel MRI-compatible brain ventricle phantom for validation of segmentation and volumetry methods. *Journal of Magnetic Resonance Imaging*, 36(2):476–482 (2012).
- [27] Fujimoto, K., Robertson, T. V., Douet, V., Garmire, D. G., and Stenger, V. A. A Structurally Anthropomorphic Brain Phantom. In *ISMRM 23th Annual Meeting & Exhibition*. Toronto, Ontario, Canada (2015).
- [28] Gopalan, K., Tamir, J., Arias, A., and Lustig, M. Toward 3D Printed, Anatomy-Mimicking, Quantitative MRI Phantoms. In *ISMRM 25th Annual Meeting & Exhibition*. Honolulu, Hawaii, USA (2017).
- [29] Reich, D. S., Lucchinetti, C. F., and Calabresi, P. A. Multiple Sclerosis. *New England Journal of Medicine*, 378(2):169–180 (2018).
- [30] Kumar, D. R., Aslinia, F., Yale, S. H., and Mazza, J. J. Jean-martin charcot: The father of neurology. *Clinical Medicine and Research*, 9(1):46–49 (2011).

- [31] Kamm, C. P., Uitdehaag, B. M., and Polman, C. H. Multiple sclerosis: Current knowledge and future outlook. *European Neurology*, 72(3-4):132–141 (2014).
- [32] Pugliatti, M., Rosati, G., Carton, H., Riise, T., Drulovic, J., Vécsei, L., and Milanov, I. The epidemiology of multiple sclerosis in Europe. *European Journal of Neurology*, 13(7):700–722 (2006).
- [33] Brownlee, W. J., Hardy, T. A., Fazekas, F., and Miller, D. H. Diagnosis of multiple sclerosis: progress and challenges. *The Lancet*, 389(10076):1336–1346 (2017).
- [34] Lublin, F. D., Reingold, S. C., Cohen, J. A., Cutter, G. R., Sørensen, P. S., Thompson, A. J., Wolinsky, J. S., Balcer, L. J., Banwell, B., Barkhof, F., Bebo, B., Calabresi, P. A., Clanet, M., Comi, G., Fox, R. J., Freedman, M. S., Goodman, A. D., Inglese, M., Kappos, L., Kieseier, B. C., Lincoln, J. A., Lubetzki, C., Miller, A. E., Montalban, X., O’Connor, P. W., Petkau, J., Pozzilli, C., Rudick, R. A., Sormani, M. P., Stüve, O., Waubant, E., Polman, C. H., and Polman, C. H. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology*, 83(3):278–86 (2014).
- [35] Stys, P. K., Zamponi, G. W., van Minnen, J., and Geurts, J. J. G. Will the real multiple sclerosis please stand up? *Nature Reviews Neuroscience*, 13(7):507–514 (2012).
- [36] Gheraldes, R., Ciccarelli, O., Barkhof, F., De Stefano, N., Enzinger, C., Filippi, M., Hofer, M., Paul, F., Preziosa, P., Rovira, A., DeLuca, G. C., Kappos, L., Yousry, T., Fazekas, F., Frederiksen, J., Gasperini, C., Sastre-Garriga, J., Evangelou, N., and Palace, J. The current role of MRI in differentiating multiple sclerosis from its imaging mimics. *Nature Reviews Neurology*, 14(4):199–213 (2018).
- [37] Tillema, J. M. and Pirko, I. Neuroradiological evaluation of demyelinating disease. *Therapeutic Advances in Neurological Disorders*, 6(4):249–268 (2013).
- [38] Kilsdonk, I. D., Jonkman, L. E., Klaver, R., Van Veluw, S. J., Zwanenburg, J. J., Kuijjer, J. P., Pouwels, P. J., Twisk, J. W., Wattjes, M. P., Luijten, P. R., Barkhof, F., and Geurts, J. J. Increased cortical grey matter lesion detection in multiple sclerosis with 7 T MRI: A post-mortem verification study. *Brain*, 139(5):1472–1481 (2016).
- [39] Kale, N. Optic neuritis as an early sign of multiple sclerosis. *Eye and brain*, 8:195–202 (2016).
- [40] Kearney, H., Miller, D. H., and Ciccarelli, O. Spinal cord MRI in multiple sclerosis—diagnostic, prognostic and clinical value. *Nature Reviews Neurology*, 11(6):327–338 (2015).
- [41] Rincon, F. and Wright, C. B. Current pathophysiological concepts in cerebral small vessel disease. *Frontiers in aging neuroscience*, 6:24 (2014).
- [42] Gorelick, P. B., Scuteri, A., Black, S. E., Decarli, C., Greenberg, S. M., Iadecola, C., Launer, L. J., Laurent, S., Lopez, O. L., Nyenhuis, D., Petersen, R. C., Schneider, J. A., Tzourio, C., Arnett, D. K., Bennett, D. A., Chui, H. C., Higashida, R. T.,

- Lindquist, R., Nilsson, P. M., Roman, G. C., Sellke, F. W., Seshadri, S., American Heart Association Stroke Council, Council on Epidemiology and Prevention, Council on Cardiovascular Nursing, Council on Cardiovascular Radiology and Intervention, and Council on Cardiovascular Surgery and Anesthesia, o. b. o. t. A. H. A. S., on Prevention, C., Epidemiology, , Nursing, C. o. C., on Intervention, C., Cardiovascular Radiology and, Anesthesia, C. o. C. S., , and Anesthesia. Vascular contributions to cognitive impairment and dementia: a statement for healthcare professionals from the american heart association/american stroke association. *Stroke*, 42(9):2672–713 (2011).
- [43] Haffner, C., Malik, R., and Dichgans, M. Genetic factors in cerebral small vessel disease and their impact on stroke and dementia. *Journal of Cerebral Blood Flow and Metabolism*, 36(1):158–171 (2016).
- [44] De Silva, T. M. and Miller, A. A. Cerebral small vessel disease: Targeting oxidative stress as a novel therapeutic strategy? *Frontiers in Pharmacology*, 7:1–18 (2016).
- [45] Yates, P. A., Villemagne, V. L., Ellis, K. A., Desmond, P. M., Masters, C. L., and Rowe, C. C. Cerebral microbleeds: a review of clinical, genetic, and neuroimaging associations. *Frontiers in Neurology*, 4:205 (2014).
- [46] Barkhof, F. The clinico-radiological paradox in multiple sclerosis revisited. *Current Opinion in Neurology*, 15(3):239–245 (2002).
- [47] Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., and Dronkers, N. F. Voxel-based lesion–symptom mapping. *Nature Neuroscience*, 6:448–450 (2003).
- [48] Rorden, C., Karnath, H.-O., and Bonilha, L. Improving lesion-symptom mapping. *Journal of Cognitive Neuroscience*, 19(7):1081–1088 (2007).
- [49] Mah, Y. H., Husain, M., Rees, G., and Nachev, P. Human brain lesion-deficit inference remapped. *Brain*, 137(9):2522–2531 (2014).
- [50] Kurtzke, J. F. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*, 33(11):1444–52 (1983).
- [51] Fischer, J. S., Rudick, R. A., Cutter, G. R., and Reingold, S. C. The Multiple Sclerosis Functional Composite measure (MSFC): an integrated approach to MS clinical outcome assessment. *Multiple Sclerosis Journal*, 5(4):244–250 (1999).
- [52] Rao, S. M., Leo, G. J., Bernardin, L., and Unverzagt, F. Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. *Neurology*, 41(5):685–91 (1991).
- [53] Makris, N., Kennedy, D. N., McInerney, S., Sorensen, A. G., Wang, R., Caviness, V. S., and Pandya, D. N. Segmentation of subcomponents within the superior longitudinal fascicle in humans: A quantitative, in vivo, DT-MRI study. *Cerebral Cortex*, 15(6):854–869 (2005).

- [54] Bermel, R. A. and Bakshi, R. The measurement and clinical relevance of brain atrophy in multiple sclerosis. *The Lancet Neurology*, 5(2):158–70 (2006).
- [55] Zivadinov, R. and Bakshi, R. Central nervous system atrophy and clinical status in multiple sclerosis. *Journal of Neuroimaging*, 14(3):27–35 (2004).
- [56] Bergsland, N., Horakova, D., Dwyer, M. G., Uher, T., Vaneckova, M., Tyblova, M., Seidl, Z., Krasensky, J., Havrdova, E., and Zivadinov, R. Gray matter atrophy patterns in multiple sclerosis: a 10-year source-based morphometry study. *NeuroImage: Clinical*, 17:444–451 (2018).
- [57] Lambert, C., Benjamin, P., Zeestraten, E., Lawrence, A. J., Barrick, T. R., and Markus, H. S. Longitudinal patterns of leukoaraiosis and brain atrophy in symptomatic small vessel disease. *Brain*, 139(4):1136–1151 (2016).
- [58] Despotović, I., Goossens, B., and Philips, W. MRI segmentation of the human brain: challenges, methods, and applications. *Computational and Mathematical Methods in Medicine*, 2015(450341):1–23 (2015).
- [59] Zhang, Y., Brady, M., and Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57 (2001).
- [60] Fischl, B. FreeSurfer. *NeuroImage*, 62(2):774–781 (2012).
- [61] Ashburner, J. and Friston, K. J. Unified segmentation. *NeuroImage*, 26(3):839–851 (2005).
- [62] Maclaren, J., Han, Z., Vos, S. B., Fischbein, N., and Bammer, R. Reliability of brain volume measurements: A test-retest dataset. *Scientific Data*, 1:1–9 (2014).
- [63] Alfano, B., Comerci, M., Larobina, M., Prinster, A., Hornak, J. P., Selvan, S. E., Amato, U., Quarantelli, M., Tedeschi, G., Brunetti, A., and Salvatore, M. An MRI digital brain phantom for validation of segmentation methods. *Medical Image Analysis*, 15(3):329–339 (2011).
- [64] Kasten, J. A., Vetterli, T., Lazeyras, F., and Van De Ville, D. 3D-printed Shepp-Logan phantom as a real-world benchmark for MRI. *Magnetic Resonance in Medicine*, 75(1):287–294 (2016).
- [65] DeWerd, L. A. and Kissick, M., editors. *The Phantoms of Medical and Health Physics*. Springer, New York (2014).
- [66] Ihalainen, T. M., Lönnroth, N. T., Peltonen, J. I., Uusi-Simola, J. K., Timonen, M. H., Kuusela, L. J., Savolainen, S. E., and Sipilä, O. E. MRI quality assurance using the ACR phantom in a multi-unit imaging center. *Acta Oncologica*, 50(6):966–972 (2011).
- [67] Gunter, J. L., Bernstein, M. A., Borowski, B. J., Ward, C. P., Britson, P. J., Felmlee, J. P., Schuff, N., Weiner, M., and Jack, C. R. Measurement of MRI scanner performance with the ADNI phantom. *Medical Physics*, 36(6):2193–2205 (2009).

- [68] Jiang, Y., Ma, D., Keenan, K. E., Stupic, K. F., Gulani, V., and Griswold, M. A. Repeatability of magnetic resonance fingerprinting T1 and T2 estimates assessed using the ISMRM/NIST MRI system phantom. *Magnetic Resonance in Medicine*, 78(4):1452–1457 (2017).
- [69] Aliaga, E. S. and Barkhof, F. MRI mimics of multiple sclerosis. *Handbook of Clinical Neurology*, 122:291–316 (2014).
- [70] Liu, Z., Pardini, M., Yaldizli, Ö., Sethi, V., Muhlert, N., Wheeler-Kingshott, C. A., Samson, R. S., Miller, D. H., and Chard, D. T. Magnetization transfer ratio measures in normal-appearing white matter show periventricular gradient abnormalities in multiple sclerosis. *Brain*, 138(5):1239–1246 (2015).
- [71] Martinez Sosa, S. and Smith, K. J. Understanding a role for hypoxia in lesion formation and location in the deep and periventricular white matter in small vessel disease and multiple sclerosis. *Clinical Science*, 131(20):2503–2524 (2017).
- [72] Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., and Mechelli, A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neuroscience & Biobehavioral Reviews*, 36(4):1140–1152 (2012).
- [73] Goodin, D. S. Magnetic resonance imaging as a surrogate outcome measure of disability in multiple sclerosis: have we been overly harsh in our assessment? *Annals of neurology*, 59(4):597–605 (2006).
- [74] Eggert, L. D., Sommer, J., Jansen, A., Kircher, T., and Konrad, C. Accuracy and reliability of automated gray matter segmentation pathways on real and simulated structural magnetic resonance images of the human brain. *PloS one*, 7(9):e45081 (2012).
- [75] Tudorascu, D. L., Karim, H. T., Marongé, J. M., Alhilali, L., Fakhraan, S., Aizenstein, H. J., Muschelli, J., and Crainiceanu, C. M. Reproducibility and bias in healthy brain segmentation: comparison of two popular neuroimaging platforms. *Frontiers in neuroscience*, 10(503):503 (2016).