Facing the issues of deep grey matter segmentation in MS

de Sitter, A.

2020

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
de Sitter, A. (2020). Facing the issues of deep grey matter segmentation in MS.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl
Introduction

MS

Multiple sclerosis (MS) is a neurological disease and, according to Dutch National MS fund, in the Netherlands 1 out of 1000 people are diagnosed with the disease. Often the diagnosis is given to individuals between 20 and 50 years old and more frequently to women than men (1). Both the symptoms and the disease course can vary substantially between individuals. Symptoms that are often experienced are abnormal sensations, pain or burning feelings in limbs, visual problems, difficulty walking, fatigue and balance problems. Recent studies also showed that cognition can be affected by MS, for example impairing memory and concentration (2, 3).

There are three main disease types in MS (4); relapsing remitting, with alternating periods of disease activity and stability; secondary progressive, phase following relapsing-remitting in majority of cases (5), with worsening of central nervous system and still disease activity with alternating periods with stable periods; primary progressive, steadily worsening symptoms from onset and without stable periods.

The exact cause of MS is unknown and over the years multiple options have been proposed; genetic causes, environmental causes or infectious agents. MS is characterized by typical focal inflammatory, demyelinating lesions and degeneration and volume loss of grey matter (GM) (6, 7). These pathological changes can be depicted in vivo using magnetic resonance imaging (MRI) (8, 9). MS lesions are typically seen in the white matter (WM) of the brain, however also occur in GM. The degeneration and consequent volume loss of GM is often referred to as GM atrophy (2, 10).
GM atrophy in MS occurs both in cortical as subcortical GM (7, 11). The atrophy in the deep GM (dGM) structures such as caudate nucleus, putamen and thalamus is associated with clinical and cognitive impairment (3, 11-15). However, that the cause and direct consequence of dGM atrophy in MS is still not known. Therefore, multiple studies have suggested that future studies should focus on dGM atrophy in MS aiming to understand the atrophic process and the impact of dGM atrophy (16, 17).

MRI

MRI plays a crucial role in diagnosis and research of MS. MRI makes in vivo identification of WM lesions possible. A combination of the amount of lesions and recurrences of the disease (e.g. more lesions or clinical relapse) over time create the diagnostic criteria for MS. These criteria have been adjusted over time with advancing knowledge (18, 19) and the guidelines also suggest preferred scan sequences. For diagnosing MS conventional imaging is often used; T1-weighted and T2-weighted, preferable fluid-attenuated inversion recovery (FLAIR) or otherwise proton density (PD). In Figure 1 an axial slice of a T1-weighted and FLAIR scan of a subject with MS is shown. On both images a few distinct WM MS lesions are marked.

On T1-weighted images the deep GM (dGM) structures can be well distinguished from the WM, shown in Figure 2. By delineations of dGM structures on the image, their volume can be measured. Moreover, when using longitudinal scans, i.e. two or more scans of one person at different times points, the difference in volume of dGM structure can be measured and if volume loss occurs we can define the amount of dGM atrophy.

More advanced MRI sequences and technics are mainly used in research. For example, functional MRI (fMRI) is used to measure changes in motor, visual and cognitive networks (20). Quantitative MRI sequences include diffusion tensor image (DTI) and Quantitative Susceptibility Mapping (QSM) to measure tissue integrity and iron content (21, 22). However, conventional techniques are also still used in MS research, for example measuring brain volumes or lesion volumes (23, 24).

In this thesis we used only conventional MRI sequences such as T1-weighted and FLAIR and PD. With these sequences we can do both research on lesions as (accurate) measurement of dGM atrophy which is the focus of this thesis.

Figure 1: An axial slice of a T1-weighted and FLAIR scan of a subject with MS is shown. On both images a few distinct white matter MS lesions are marked. FLAIR = fluid-attenuated inversion recovery

Figure 2: An axial slice of T1-weighted image with both the deep grey matter (caudate nucleus in blue, putamen in red and thalamus in green) and white matter marked.
Open science and patient privacy

Over the years, open science has been encouraged in research fields and multiple journals, as open science could have many benefits to research. Sharing patient data (demographic and clinical data, MRI) can make meta and mega-analysis possible, provides a basis for methodological aspects of studies and collaboration between groups can be encouraged. However, there are ethical issues on open science around patients data. For example, in neuroradiological research around sharing participant MR data. The identifying information such as name, date of birth, or any national or hospital-based registration numbers saved in the metadata of the MRI scans can be removed before sharing. However, because of sufficient skin to air contrast and spatial resolution in typical structural MRI, facial recognition from a 3D-rendered version of the image is possible (25-29). Therefore, in order to share MR data, more than a simply data-transfer agreement and anonymization might be needed. For example, it has been suggested that facial features should also be removed (30-33).

Aim and overview of this thesis

The aim of this study is to improve the measurement of dGM atrophy in MS. Absence of reliable methods for accurate quantification of (d)GM atrophy in MS is a major hurdle preventing the implementation of single-patient measures of (d)GM atrophy in a clinical setting. To understand the challenges we could face during this thesis we first started, in Chapter 1, with a literature review on current hurdles in measuring GM atrophy in MS. The literature review showed that improvement of automated segmentation methods is needed for accurate measuring of GM atrophy. We expect this also to be true for dGM structures such as caudate nucleus, putamen and thalamus. Therefore, in Chapter 2, we investigate the performance of existing automated segmentation software packages in terms of the accuracy of their dGM segmentations, and compare the performance on MS subjects and healthy controls. Moreover, we studied the relation of the performance of automated segmentation software with WM lesions and GM atrophy.

Because the WM lesions play a key role in MS and could have an effect on the measurement of dGM atrophy in MS (10, 34-36), we address the segmentation of WM lesions in Chapter 3. Firstly, in Chapter 3.1 we investigate the performance of different automated WM lesions segmentation software packages. Secondly, in Chapter 3.2 we introduce a software tool to simulate lesion from MS subjects into healthy controls. With this tool we hope that future studies can investigate in more details the effect of lesions on (d)GM atrophy or other interesting issues.
English Summary
The aim of this study was to improve the measurement of deep grey matter (dGM) atrophy in multiple sclerosis (MS) with the use of magnetic resonance imaging (MRI). For this we started with discussing the challenges of measuring GM in MS in Chapter 1. One of the conclusions was that more accurate automated segmentations methods are needed. Therefore, in Chapter 2, we evaluated the relation of the performance of well-established automated segmentation software with MS pathology. In Chapter 3 we addressed the white matter (WM) lesions, as WM lesions play an important role in diagnosis of MS and also affect brain image analyses in MS. To further stimulate methodological improvements in measurement of dGM atrophy in MS, we discussed in Chapter 4 how to improve open science in this field. Lastly, in Chapter 5 we developed an MS-specific automated segmentation software, MS-SMART, and an open reference dataset.

In Chapter 1 we discussed the urgent challenges of measuring grey matter (GM) atrophy in MS, distinguishing two main fields; i. pathology, physiology, and treatment effects and ii. measurement challenges. We discussed in more detail the pathological substrate, evolution of GM atrophy, influence of physiological variability, and evaluation of treatment response. Regarding technical measurement challenges, we discussed the influence of WM lesions on the measurement of GM atrophy. Moreover the influence of atrophy itself, the influence of other MS pathology, and technical variability. For every discussion point, we provided specific recommendations (summarized in Box 1 of Chapter 1) to improve measurements and interpretation of GM atrophy in individual MS patients. Two of these recommendations provide the basis for the rest of this thesis; the need of a public available reference data set and improvement of segmentation methods for MS.

In Chapter 2, we investigated the performance of existing automated dGM segmentation methods compared to a manual reference. Moreover, we evaluated whether there was a relation of the performance of those automated dGM segmentation methods with WM lesions and GM volume. We evaluated four different automated segmentation methods (FSL-FIRST, FreeSurfer, GIF and volBrain) on a multi-center dataset (21 MS subjects and 11 healthy controls). The performance of the methods was evaluated to manual reference on both volumetric (intraclass correlation (ICC)) and spatial (dice similarity coefficient (DSC)) agreement. The relation between segmentation accuracy of the methods, as expressed by their DSC with the manual outlines, and the global and local lesion volumes, region of interest volume, and normalized brain volume, was assessed. We concluded that existing automated methods have impaired performance on data of MS subjects, specifically, that the accuracy of the segmentations is reduced. Moreover, it was observed that performance generally deteriorated with higher lesion volume, and with lower normalized brain volume and structure of interest volume. This suggests that MS pathology may contribute to the impaired performance.

In Chapter 3 we investigated several aspects of the WM lesions in MS, divided over two sub-chapters. In Chapter 3.1 we discussed the performance of five automated WM lesion segmentation methods on a multi-center MS dataset (70 MS subjects). On the 2D fluid attenuated inversion recovery (FLAIR) images, manual lesion segmentation was performed and the segmentations of five automated methods (Cascade, LST-LGA, LST-LPA, Lesion TOADS and kNN-TTP) were compared to the manual outlines. Both volumetric (ICC) and spatial agreements (DSC and false positive and false negative volumes) were assessed. Furthermore, analyses were repeated using a leave-one-center-out design to exclude the center of interest from the training phase, in order to evaluate the performance of the method on ‘unseen’ centers. We concluded that the performance of the methods in this multi-center MS dataset was moderate, but appeared to be robust even with new datasets from centers not included in training the automated methods.

In Chapter 3.2 we developed a lesion simulation method (LESIM) to improve objective investigations of the effects of WM lesions on image analyses methods and to facilitate the development of segmentation methods that are robust to the presence of WM lesions. The LESIM software simulates lesions from an MS patient into a 3D T1-weighted (3DT1) image of a healthy control (HC), which results in a modified HC 3DT1 image with realistic lesions. We evaluated LESIM by visual inspection as well as a quantitative analysis of the effect of simulated lesions on FSL-SIENAX GM segmentations. We concluded that LESIM is a new, robust, and flexible tool for reliable WM MS lesion simulation that produces realistic lesions in healthy control images. Moreover, we showed that the simulated WM lesions have the expected effect on GM segmentation using FSL-SIENAX.

In Chapter 4 we addressed the issue of open science in the field of neuro-radiology. Firstly, by assessing the impact of facial features removal on clinically relevant outcome measurements (Chapter 4.1) and secondly, by developing and evaluating a standardized protocol for manual delineations of dGM structures (Chapter 4.2).

So, in Chapter 4.1 we investigated if removing facial features would affect subsequent automated image analyses. To do so, we tested the effect of three facial features removal methods (QuickShear, FaceMasking, and Defacing) on automated image analyses methods that give clinically relevant outcome measurements. We used three datasets of different diseases: Alzheimer’s Disease, MS, and patients with a glioblastoma. Therefore, we also used three different clinically relevant outcome measurements, respectively, normalized brain volume, white matter lesion volume and tumor volume. Differences between outcomes
In Chapter 4.2 we discussed the development and evaluation of a manual segmentation protocol of dGM structures in MS. Next, we evaluated the accuracy of FASTSURF, a semi-automated segmentation method, in which sparse delineations serve as input. The standardized protocol was specifically developed for manually tracing dGM structures on 3D T1-weighted MRI scans of MS patients, by neurologists and neuroradiologists with broad experience in the field of MS and MRI. Anatomical definitions were specified for each structure and alongside these landmarks, strict guidelines on how to recognize the outermost edges of the structures on orthogonal planes were described.

To evaluate the protocol, three raters delineated dGM structures bilaterally on 3D-T1-weighted multi-center MRI scans of 23 MS patients and 12 controls. Intra- and inter-rater agreements were assessed through volumetric (ICC) and spatial (JI and CIgen) agreement. Segmentations made with FASTSURF were also evaluated in terms of both volumetric and spatial agreement. We showed that raters achieved good to excellent intra- and inter-rater agreement and that these agreements were similar with use of FASTSURF. We concluded that the dGM manual segmentation protocol showed good reproducibility within and among raters. Moreover, this protocol could be combined with FASTSURF to produce a reference set of dGM structures with a lower workload.

In Chapter 5 we discussed the development of an MS-specific dGM automated segmentation method. MS-SMART is an open source automated segmentation method and is an atlas-based approach. The atlases for MS-SMART were manual outlined on 120 (100 MS subjects and 20 healthy controls) T1 MR images with use of the protocol developed in Chapter 4.2. The use of MS-specific atlases (images and labels) could help reduce the influence of MS pathology during alignment of the atlases to the target (input) image. In total, 60 images were used as an? atlas and training set for SMART and the other 60 were used for the evaluation of SMART and two well-established automated segmentation methods (FSL-FIRST and FreeSurfer). Evaluation was performed on both volumetric (ICC) and spatial (DSC) agreement compared to the manual outlines. We concluded that SMART outperformed the two well-established methods on this MS data set. However, we expect that with use of the shared atlas set and software code of SMART more methodological improvements in segmentation of dGM structures in MS could be made.

De oorzaak van MS is nog niet bekend maar wel is duidelijk dat MS wordt gekenmerkt door focale inflammatoire, demyeliniserende laesies in voornamelijk de witte stof (WS) en degeneratie en volumeverlies van grijze stof (GS) in het centrale zenuwstelsel (de hersenen en ruggenmerg) (6, 7). Deze veranderingen kunnen in vivo zichtbaar worden gemaakt door gebruik van Magnetic Resonance Imaging (MRI) (8, 9).

De degeneratie en volumeverlies van de GS wordt vaak GS atrofie (krimping) genoemd en kan bij MS zowel corticaal als subcorticaal plaatsvinden (2, 7, 10, 11). De subcorticale atrofie treedt op in de diepe GS (dGS) structuren zoals nucleus caudatus, putamen en thalamus (3, 11-15). De precieze oorzaak en gevolgen van atrofie van de dGS is helaas niet bekend en daarom stellen verschillende onderzoeksgroepen voor dat er meer studies moeten komen naar de oorzaak en de impact van dGS atrofie bij MS (16, 17).