Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Circulation and Medical Imaging

Peter Low

Spinal Cord Lesion Detection in Multiple Sclerosis. Comparison of Magnetic Resonance Weightings.

Master's thesis in Medical Magnetic Resonance Imaging
Supervisor: Beathe Sitter
December 2019

NTNU
Norwegian University of Science and Technology

REGION ZEALAND
ZEALAND UNIVERSITY HOSPITAL

Abstract

Various types of magnetic resonance image weightings are recommended for the detection of multiple sclerosis lesions in the spinal cord. There are varying levels of diagnostic efficacy between them, due to differences in contrast ratios. Short-tau inversion-recovery (STIR) has good contrast, but has historically been viewed as lacking specificity due to imaging artefacts and noise. Phase-sensitive inversion-recovery (PSIR) has shown potential for greater contrast at 3 teslas, but has not been investigated at 1.5 tesla. Higher resolution scans, particularly three-dimensional scans can possibly detect more lesions. A study was conducted in order to investigate relative diagnostic efficacies at 1.5 tesla for the whole spinal cord for four image-weightings.

The following sequence weightings were included in this study: T2, PSIR, two-dimensional short-tau inversion-recovery (2D-STIR) and three-dimensional short-tau inversion-recovery (3D-STIR). Each weighting was scanned over two areas of anatomy – the cervical and thoracic spinal cord. Scan times of roughly four minutes and area coverage were the same for all weightings. Resolutions were varied. Two radiologists, one of consultant level, one of departmental level, separately evaluated each sequence weighting alone. Detection of lesions was recorded by level in the spinal column, and subjective evaluations of artefacts and noise were also measured. Kappa values for agreement between radiologists were calculated for each weighting. Subsequently, both radiologists consensus-reviewed all patients with the use of all images available including axial reconstructions of the 3D-STIR acquisition. This reference standard was used to calculate sensitivities and positive-predictive-values (PPV). Contrast-to-noise measurements (CNR) were done for consensus confirmed lesions by the use of regions-of-interest, which were of identical size and position between weightings.

2D-STIR had the highest sensitivity and detected more lesions than all other sequences, though there may be issues with image-based selection bias and the imperfect reference standard. This difference was statistically significant for both radiologists on cervical coverage scans and for one radiologist on thoracic scans. 2D-STIR also had significantly greater CNR than all other sequences for both areas of coverage. PPV and agreement for this sequence were high and comparable with that of T2. Diagnosis was no more affected by artefacts and noise on 2D-STIR than T2. PSIR was the least sensitive sequence, despite having a slightly better CNR than T2. 3D-STIR added a greatest number of lesions when added to 2D-STIR findings, than PSIR or T2. These additional lesions were smaller in size than the overall average. But 3D-STIR was badly affected by noise and possibly concomitant field artefacts. PPV was lowest for 3D-STIR.

2D-STIR is no more affected by artefacts than T2 and has good PPV and agreement. This sequence is now perhaps good enough to be used alone. PSIR had poor sensitivity especially on thoracic scans and CNR was significantly worse than 2D-STIR. 3D-STIR was very badly affected by noise and imaging artefacts, but can perhaps serve as a good secondary sequence for cervical scanning if examination times allow. T2 is also a reliable choice as a secondary sequence and is just as useful for thoracic scanning as for cervical.

Abstrakt

Adskillige typer for magnetisk resonans billedvægtninger er anbefalet til den bedst mulige fremstilling af multipel sklerose læsioner i rygmarven. Den diagnostiske virkningsfuldhed varierer mellem dem, grundet forskelle i kontrast forhold. Short-tau inversion-recovery (STIR) har god kontrast, men har historisk set syntes at mangle specificitet grundet billedartefakter og støj. Phase-Sensitive Inversion-recovery (PSIR) har vist god potential for bedre kontrast på 3 tesla scannere, men er ikke undersøgt på 1.5 tesla. Højere opløsning på scanninger, særligt tre-dimensionelle scanninger

kan muligvis føre til opdagelsen af flere læsioner. Et studie blev gennemført for at undersøg de relative diagnostisk virkningsfuldhed på 1.5 teslaer, ved scanning af hele rygmarven for fire vægtninger.

Følgende sekvensvægtninger blev inkluderet i studiet: T2, PSIR, to-dimensionel STIR (2D-STIR) og tre-dimensionel STIR (3D-STIR). Hver vægtning blev scannet over to anatomiske områder – den cervikale og thorakale rygmarv. Scantider på cirka fire minutter, samt anatomisk dækning af billeder var ens for alle vægtninger. Opløsning variererede mellem vægtningerne. To radiologer, en på overlæge niveau og en på afdelingsniveau, adskilleligt evalueret på hver vægtninger, hver for sig. Opdagelse af læsioner blev noteret efter højde-position i columna vertebralis og subjektive vurderinger af artefakter og støj blev også indsamlet. Kappa værdier for enighed blev beregnet mellem radiologerne for hver vægtning. Efterfølgende samledes radiologerne til et konsensus review af alle patienter med alle vægtninger, samt axiale rekonstruktioner af 3D-STIR sekvenserne. Denne referens standard anvendtes til beregning af sensitivitet, og positiv prædiktiv værdi (PPV). Kontrast-til-støj forhold (CNR) målinger udførtes på læsioner angivet på referens standarden, ved brug af regions-of-interest, som var ens i størrelse og placering mellem vægtningerne.

2D-STIR ydede den højeste sensitivitet og opdagede flere læsioner end alle andre vægtninger, dog kan der være tale om billed-baseret selektions-bias, samt den ikke-perfekte referens standard. Forskelle i antal opdagede læsioner var statistisk signifikant for begge radiologer på cervikale scanninger og for en radiolog på thorakale. 2D-STIR ydede også en signifikant bedre CNR end alle andre vægtninger. PPV og enighed for denne sekvenstype var høje og sammenlignelig med T2-vægtningen. Diagnosticerbarheden blev ikke forstyrret af artefakter eller støj i større grad på 2D-STIR end på T2. PSIR var den mindst sensitive vægtning, trods at CNR var bedre end på T2. 3D-STIR tilføjet flere læsioner end T2 eller PSIR, når samlet med 2D-STIR fund. Disse ekstra læsioner var mindre i størrelse end den overordnet gennemsnit. Dog var 3D-STIR svært påvirket af støj og muligvis også samtidig-felts artefakter. PPV og enighed var lavest for 3D-STIR, .

2D-STIR er ikke længere mere påvirkelig overfor artefakter end T2 og har god PPV samt enighed mellem radiologer. Denne sekvens er måske nu pålidelig nok til at blive anvendt alene. PSIR havde betydeligt lav sensitivitet, særligt på thorakale scanninger og CNR var signifikant værre end på 2D-STIR. 3D-STIR var slemt påvirket af støj og billedartefakter, men kan muligvis tjene godt som anden sekvens ved cervikale scanninger, hvis undersøgelsestider tillader. T2 er også et pålideligt valg som anden sekvens og denne er lige anvendelig til thorakale såvel cervikale scanning.

List of Acronyms

2D	Two-Dimensional
3D	Three-Dimensional
CMSC	Consortium of Multiple Sclerosis Centres
CNR	Contrast-to-Noise Ratio
CSF	Cerebro-Spinal Fluid
FOV	Field of View
FSE	Fast Spin Echo
IES	Inter-Echo Spacing
MAGNIMS	Magnetic Resonance in Multiple Sclerosis
MPRAGE	Magnetization Prepared Rapid Gradient Echo
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
	CIS Clinically Isolated Syndrome
	PPMS Primary Progressive MS
	RRMS Relapse-Remitting MS
	SPMS Secondary Progressive MS
NASC	Normal Appearing Spinal Cord
PD	Proton Density
PPV	Positive Predictive Value
PSIR	Phase Sensitive Inversion Recovery
ROI	Region of Interest
RIS	Roentgen Information System
SD	Standard Deviation
SE	Standard Error
STIR	Short Tau Inversion Recovery
VGA	Visual Grading Analysis

Table of Contents

Background

Multiple Sclerosis (MS)

MS is a chronic inflammatory demyelinating neurological disease of the nervous system. MS is the most widespread neurological disease affecting the central nervous system (CNS) amongst young adults in the western world. MS is incurable. Whilst the clinical course for an individual can be difficult to predict, most patients will experience chronic disability, which only progresses over time. MS causes the greatest level of disability in young people, that is not due to trauma (1).

MS is divided into subtypes, based on the observed clinical course and findings on MRI and cerebrospinal fluid (CSF) tests. These types are:

- Relapsing remitting MS (RRMS), where clear MS symptoms appear, such as reduced motor function, and then dissipate after a few days or weeks. These episodes of disease activity are known as attacks and impairments can be very debilitating. This group comprises 85% of all MS cases.
- Primary progressive MS (PPMS), where there is a gradual increase in the effect of the disease over periods of months or years, without attacks.
- Secondary progressive MS (SPMS), where there is a gradual worsening of the condition over time, with episodic periods of disease activity. RRMS can develop into SPMS after some years.
- Clinically isolated syndrome (CIS), after a single attack, where the diagnosis of MS has not definitely been made but with likely signs of MS on an MRI scan and / or analysis.

Symptoms

Reduced sensibility is a common occurrence. Sensing vibration and awareness of joint position (proprioception) becomes more difficult. Reduction in pain sensibility and the ability to feel light touches is also a common occurrence. Motor impairments, due to impairment of the function of the corticospinal tract, is a common presenting symptom of MS. This results in weakness in legs or arms, affecting most notably the ability to walk. Reduced strength in the upper extremities is also common. These symptoms can disappear partially or completely in relapsing MS or can progress steadily over time in progressive MS. Urinary problems, including incontinence can occur, both as a result of sensory and motor nerve dysfunction. Cognitive problems can also occur (2).

MS Pathophysiology in the CNS

MS is perhaps most clearly defined by the lesions it causes in the brain and spinal cord. These are commonly called plaques but are referred to as lesions in this paper. These lesions are areas where demyelination, inflammation and axonal loss have occurred (3). Demyelination is thought to be due to an auto-immune reaction.

In the spinal cord, there is evidence of inflammation, demyelination and axonal loss. There may also be gliosis, which is a reaction to damage on the spinal cord's supportive cells (4).

Myelin is a form of lipid contained in oligodendrocyte cells, which surround axons in nerve fibres. Its function is to electrically isolate the axons. In normal function, electrical signals are conducted along the axon carrying information through the nervous system. When demyelination occurs, the electrical current cannot be contained within the axon and nerve signals are interrupted. This can greatly reduce function in the nerve fibre. In addition, as a part of a likely auto-immune reaction, immune cells release cytotoxic chemicals, both within the plaque and into the surrounding region,

further affecting nerve function. Disease activity also leads to axonal loss, associated with prolonged demyelination.

Where and when a new lesion appears, disease activity can be demonstrated as contrast enhancement on MRI scans. Contrast enhancement due to active MS disease can be seen in both the brain and spinal cord (1). Examples of lesions as seen on MRI scans are shown in Figure 1.

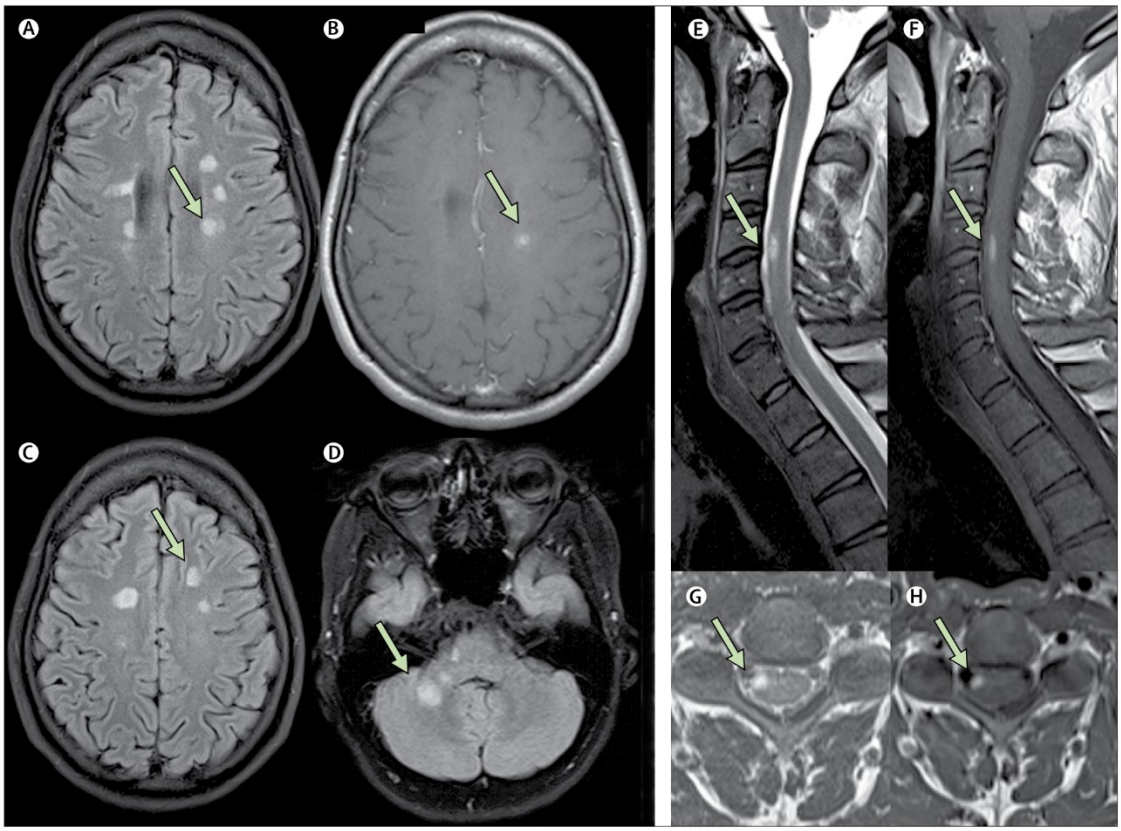

Figure 1: MRI images showing features and location of MS-lesions in the cerebrum and spinal cord. A, C and D are Fluid Attenuated T2 images showing periventricular, juxtacortical and infratentorial lesions respectively (arrows). B shows a T1 weighted image after contrast administration, where a single periventricular lesion exhibits contrast enhancement. E and F show sagittal images of the spinal cord on T2 and T1 (post-contrast) images. Arrows here point to a single anterior/lateral spinal cord lesion. There is contrast enhancement in this lesion, as seen in F. G and H show axial scans, also T2 and T1-post contrast, showing the lesion as being located in the right anterolateral area. Figure taken from Brownlee, Hardy (5).

That MS disease leads to signal changes on MRI has been pathologically confirmed in post-mortem magnetic resonance imaging (MRI) studies of deceased MS patients. These studies have confirmed that demyelination, inflammation, axonal loss and gliosis can be visualised with MRI (6-8).

Diagnostic Criteria and MRI imaging criteria for MS

There is no singular test or clinical feature for MS. Diagnosis is established by the use of diagnostic criteria, which involve various factors. Here there is an important role for MRI. Lesion detection is its primary role in MS diagnosis in conjunction with the use of diagnostic and imaging criteria.

MS can be diagnosed on the basis of clinical information provided by the patient; for example if the patient reports having had symptoms clearly indicating MS disease on two separate occasions, this fulfils the diagnostic criteria without the need for MRI. However patients often present themselves to clinicians without this clearly being the case. Often it can be the first symptoms of the disease,

precluding previous episodes. Furthermore, MS lesions do not always lead to symptoms. Notably, 30% of patients with CIS have clinically silent lesions in the spinal cord (2). In these situations, MRI plays a central role in establishing the diagnosis via the application of imaging criteria for the diagnosis of MS. Therefore, lesions detected on MRI are an important aspect of diagnosis.

There are MS-typical locations for lesions in the CNS detectable on MRI. These locations are used to diagnose MS by way of imaging criteria, known as the MacDonald Criteria, first published in 2001 by the International Panel on MS Diagnosis (9). Revisions to these have been proposed in the years since to make them more effective. In 2016 the Magnetic Resonance Imaging in MS group (MAGNIMS) published recommendations for modifications to the criteria (10) and the international panel also published revised recommendations in 2017 (11).

Dissemination in space and time in Imaging

The diagnostic criteria rely on demonstrating two key features of disease course to establish a diagnosis. They are dissemination in space and time. Both are required in order to establish MS as the diagnosis. That MS-typical symptoms and imaging features can be related to different parts of the CNS (dissemination in space - DIS) and with disease activity occurring at different points in time (dissemination in time - DIT) is the central concept in the criteria.

According to MAGNIMS' recommendations, dissemination in space can be demonstrated on MRI by the detection of two of the following (10):

- Three or more periventricular lesions
- One or more infratentorial lesion
- One or more spinal cord lesion
- One or more optic nerve lesion
- One or more cortical or juxtacortical lesions

Dissemination in time can be established on MRI in two ways. The first is the presence of two MS-typical lesions, where only one lesion exhibits contrast enhancement. This is more readily demonstrated in the brain than the spinal cord (12). The second way is the detection of a new MS-lesion on a follow-up scan conducted a reasonable period after the initial scan – for example four weeks. A spinal cord lesion on follow-up could constitute such a finding (12).

An MRI scan of the spinal cord can therefore contribute to establishing DIS and DIT. Lesion detection in spinal cord MRI is of clinical importance as the ability to do so has a direct impact on diagnosis, prognosis and treatment monitoring (12, 13). This underlines the importance of lesion detection in spinal cord MRI.

Spinal Cord Anatomy

The spinal cord is a lengthy structure, 40-45 cm long in adults, surrounded by cerebrospinal fluid, within the spinal canal of the vertebral column. It is variable in width, between one and two centimetres, being broadest in the cervical area and thinnest in the mid-thoracic area. It extends from and is a continuation of the brain stem in the cranium down to the level of the 1st or 2nd lumbal vertebra. Here it terminates at the conus medullaris.

The spinal cord consists of a central H-shaped core of grey matter almost completely surrounded by white matter. In cross section, spinal cord grey matter takes the form symmetrical horns anteriorly (cornu anterius) and posteriorly (cornu posterius). In the thoracic medulla, smaller lateral horns are found centrally on each side (cornu lateralis). The grey structure can be said to divide the white matter of the medulla into three parts or columns, the anterior, lateral and posterior cords

(respectively funiculus anterior, lateralis and posterior). These structures run the whole length of the spinal cord, forming their respective columnae. There is also a central canal (canalis centralis), which is a CSF filled structure.

Spinal Cord Lesions

Spinal cord lesions more readily lead to symptoms than cerebral lesions. More than nine out of ten MS patients will at some point acquire a spinal cord lesion. Of patients presenting with MS-like symptomology, but without symptoms of transverse myelitis, 30% will have spinal cord lesions. Spinal cord scanning for newly diagnosed or suspected MS sufferers has prognostic value (13). Lesions are rarely longer than two vertebral segments in length in the cranio-caudal direction (1-3 cm) and are most readily detected in the cervical spinal cord (1, 11).

Symptoms often reflect the lesion's position in the CNS. For instance, a left lateral cervical lesion around the C4-C6 level may give weakness in the left arm, whereas a right sided posterior lesion in the low thoracic spinal cord may lead to a lack of sense of touch and reduced proprioception in the right leg (2). Lateral lesions have been shown to be associated with limb weakness (14).

Clinical Spinal Cord MRI Image Quality in MS

Image quality in clinical MRI imaging is primarily determined by four factors: contrast-to-noise ratio (CNR), spatial resolution, scan time and signal-to-noise ratio (SNR). CNR is a primary factor in determining detectability of lesions and is related to SNR (15). There is a trade-off between SNR, resolution and short scan times. For example, higher resolution will reduce SNR if scan times are kept constant. This can potentially affect CNR, so that lesions become less apparent. Alternatively, greater SNR and therefore perhaps better CNR, require reduced resolution, so that smaller lesions may be missed (16).

Clinical MRI does not detect all MS lesions in the spinal cord (7, 17). This is partly because the spinal cord is a rather challenging anatomical structure to image. It is a long and thin structure, where there is a need for an overview over a large area, as it is typically 40-50 centimetres in length, but simultaneously a need for relatively high resolution in order to image relatively small lesions. These may be on the scale of a few millimetres. Given that scan times may not be too long to be clinically acceptable, resolution is limited if a reasonable SNR is to be maintained. Post-mortem examinations with very high resolution MRI at 4.7T of MS-patients show that clinical imaging does not detect all lesions that are present (6, 7, 17).

Artefacts also affect spinal cord imaging. Reduced resolution can lead to adverse effects from truncation artefacts and partial volume averaging. These can easily obscure pathology, especially small lesions. Physiological patient factors, such as breathing, pulsation of cerebrospinal fluid (CSF), blood vessel pulsation, swallowing and gross patient movement also reduce image quality and diagnostic accuracy (18).

There are many different types of sequence and various weightings in MRI. Each has its own characteristics in terms of inherent signal, scan time, level of artefacts and pathological contrast and each yields its own trade-offs between these factors. An important example is the short-tau-inversion -recovery (STIR) sequence which whilst sensitive to lesions, has historically been very susceptible to flow and pulsation artefacts, limiting its usefulness (18). Additionally, some sequences have more inherent signal than others. This is important in weighing the trade-off between signal, resolution and scan time when deciding which sequence to use in clinical practice. The large number of options and the different trade-offs each choice of sequence represents can be difficult to surmise.

Recommended Sequences for MS Spinal Cord imaging

International panels, MAGNIMS and the Consortium of MS Centers (CMSC) Task Force, have published guidelines for choice of sequences in spinal cord imaging in MS (19, 20). International Panels' recommendations are based on reviews of peer-reviewed research. Both sets of recommendations advise the use of at least two different recommended sequence types. This is because research indicates that some sequences, such as STIR, are more sensitive, but are less specific, whereas others, such as T2, are more specific, but less sensitive. The two sequences can therefore supplement each other in the task of lesion detection. The more sensitive, but less specific sequence can detect more possible lesions, which can then be confirmed as actual lesions or not on the more specific, but less sensitive sequence.

The following sequences are recommended by CSMC:

- sagittal T2
- sagittal STIR
- sagittal Proton-weighted (PD)

MAGNIMS recommends the following are used:

- Dual-echo (PD and T2)
- STIR (instead of PD)

As an option, sagittal T1-weighted phase-sensitive-inversion-recovery (PSIR) is also recommended as a possible alternative to STIR for cervical scans. Post contrast sagittal T1 scans should be done if lesions are present. Another option named is to scan the cervical spinal cord axially in its entirety. For axial imaging, CMSC recommends this is done routinely for lesions detected on sagittal images. MAGNIMS lists axial 2D or 3D T2 weighted FSE as optional sequences.

Recommended parameters for sagittal imaging are slices no thicker than 3mm, with in plane resolution no greater than $1mm^2$. Three dimensional imaging is not included in the list of recommended sagittal sequences.

Technical description of sequences

T2-weighted Two-dimensional Fast-Spin-Echo imaging (2D-FSE)

T2 weighted 2D-FSE is a very standard imaging sequence in neurological MRI. It is characterized by high signal from body fluids, most notably CSF. Oedema caused by various pathologies including MS lesions increases the T2 times for neural tissue, such that these appear with increased signal, as

Figure 2 shows a sagittal T2-weighted FSE image of the spinal cord. The spinal cord itself is seen as a dark grey structure surrounded by cerebrospinal fluid and intraspinal fat.

discussed above. Because of its high signal, relatively low level of artefacts and good specificity, it is a sequence type very often used in neurological MRI. An example is shown in Figure 2.

Typical parameters for a T2 weighted sequence are a TR over 2000ms and a TE between 80 and 120 ms (15). This sequence produces good signal, as both water and fat contribute. Fat's contribution arises from the use of many echoes in an echo train, which is a feature of the FSE technique. When many echoes are collected, some echoes will be measured rather soon after excitation and these echoes will contain more signal from fat. Overall T2 contrast is produced by measuring large spatial frequencies at echo-times corresponding to the desired image contrast. The desired echo time can be referred to as the effective echo time. This is most often the quantity controlled by the TE value input on an MRI scanner (15).

As T2 weighted sequences have a relatively high inherent signal, it is generally possible to increase resolution relative to other sequences, such as STIR, without the need for increasing acquisition time to compensate. Higher resolution may be useful for the detection of smaller MS lesions.

Inversion Recovery

Inversion recovery (IR) is a technique with which the T1-contrast of an MRI image can be manipulated. It can be used in one of two ways:

- to selectively suppress signal from tissues of a particular T1 recovery time
- to increase T1 contrast between tissues relative to a conventional T1-weighted image.

Overall, an IR pulse sequence can be divided into two parts; a radio-frequency (RF) inversion pulse and a self-contained pulse sequence module, which can be referred to as the host sequence (21). As such, inversion pulses can be used in a variety of sequences.

The repetition cycle of the pulse sequence begins with the RF pulse. This inverts the longitudinal magnetization (Mz): it is flipped by a 180 degree radio-frequency pulse such that it points in the negative z-direction. It subsequently recovers without developing transverse magnetisation, going through a null point where its value is zero. The inversion pulse can be either slice-selective or non-selective, whereby the entire scan volume is inverted. The degree to which tissue magnetization is flipped can vary slightly spatially. Differences can occur due to static field or radio field inhomogeneity, or variations in the degree of RF manipulation in the slice profile with slice-selective inversion. As this would otherwise interfere with subsequent imaging, giving rise to artefacts, inversion pulses are usually followed immediately by spoiler gradients. These eliminate any unwanted transverse magnetization. The recovery of magnetizations after inversion is shown in Figure 3.

After a specific amount of time, an excitation pulse of 90 degrees is applied. The amount of time between inversion and excitation defines the imaging parameter called inversion time. This parameter is of crucial importance with regards to the weighting of the image. Two or three inversion pulses can be chained together in order to suppress signal from two or three tissue types, as in double inversion recovery for example. Inversion can also be spectrally selective, such that only lipids are inverted.

Excitation marks the start of the host sequence. Whatever amount of Mz is available at this point, will be flipped into the transverse plane, where it can be measured for imaging. This is the case regardless of whether the magnetization was in the positive or negative z-direction upon excitation. The host sequence can consist of a variety of different pulse sequence types, including FSE

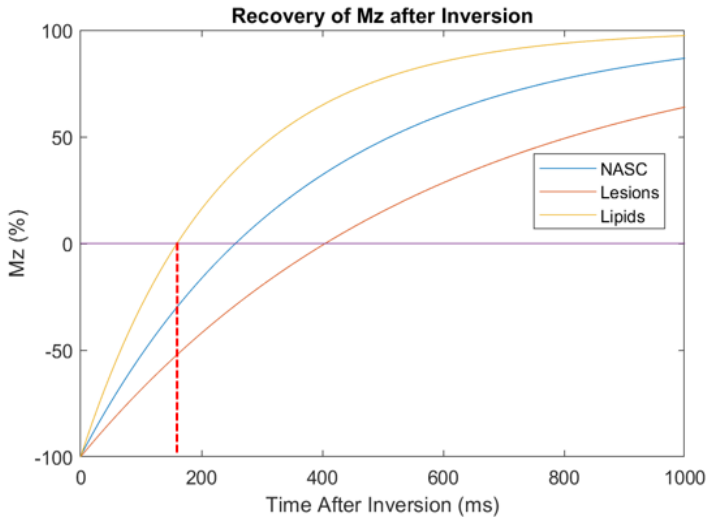

Figure 3: Shows the T1-recovery curves of lipids, normal spinal cord tissue (NASC), and MS-lesion tissue. The T1-values are adapted from Bot, 2004. The dashed line shows the null point in time after inversion, where longitudinal lipid magnetization is zero.

sequences and gradient-echo sequences such as magnetization prepared rapid gradient echo (MPRAGE) sequences. There are many options, as inversion recovery can be used in conjunction with many other MRI techniques.

Short Tau Inversion Recovery (STIR)

This sequence is among those recommended for spinal cord imaging in MS. In STIR imaging, a short inversion time is used to selectively suppress the signal from lipids. STIR images have bright CSF, a dark grey spinal cord and dark fat. Oedema due to pathology, including MS-lesions, appear as bright spots in the spinal cord. An example of 2D STIR is shown in Figure 4.

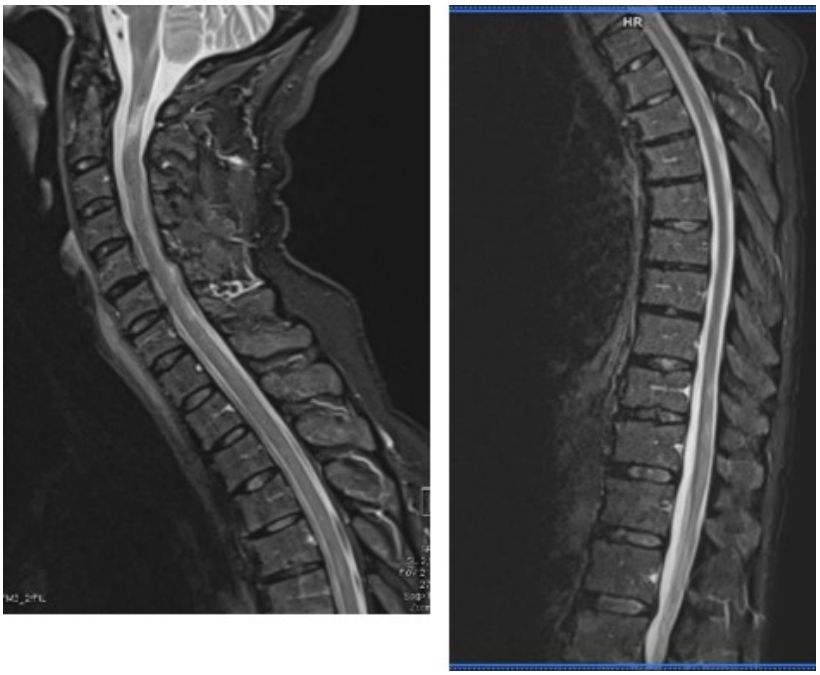

Figure 4: shows 2D-STIR images of the spinal cord. CSF is bright, fat is dark and the spinal cord is dark grey. An MS-lesion can be seen at the C2 level and in the lower thoracic spinal cord. The TE value of this image is 38 ms.

Because the CSF is so bright, flow and pulsation in the spinal canal can cause significant artefacts. The sequence is also prone to Gibb's or truncation artefacts for similar reason. The large and abrupt change over a small distance from bright CSF to dark grey spinal cord in the image can easily be above the highest spatial frequency sampled by the scanner, causing the artefact.

The appropriate inversion time to suppress a tissue of given T1-recovery time is defined as:

$$Inversion\ Time_{Tissue\ A} = ln2 * T1\ recovery\ time_{Tissue\ A}$$

Eq. 1: suppression inversion time.

Lipids generally have a short T1 time of around 2-300 milli-seconds, making an inversion time of roughly 160 milliseconds useful. Setting the parameter to this value ensures that signal from fat reaches null point and has no longitudinal magnetization at the point of excitation. As a consequence, no signal is acquired from lipids. This is shown in Figure 3.

The repetition time of the sequence should be longer than 3 seconds, so that lesions' magnetization is close to fully recovered. This contributes to a greater signal from lesions. However, if the TR is too long, for instance above 5 seconds, then the CSF can become too bright and can begin to obscure low contrast details in the spinal cord including lesions. A medium value of 3.8 seconds could be regarded as a reasonable value. Longer TE times reduce signal due to T2 decay whilst very short TE times remove all T2 weighting. An appropriate TE could therefore be 38 milli-seconds, which largely corresponds to a T2-type weighting, but with shorter than typical values to improve signal.

Phase Sensitive Inversion Recovery (PSIR)

PSIR images as used in MS imaging can be thought of as being T1 weighted. They have bright fat, very dark CSF and a mid grey background. MS lesions appear as dark spots. As the CSF is dark, flow artefacts are theoretically suppressed. An example PSIR-image is shown in Figure 6.

As well as lengthening T2 times, MS-lesions also lengthen T1 times. This means that there is potential to exploit this difference in order to create diagnostic image contrast. In neuroimaging, T1 contrast can be accentuated by using inversion pulses and combining this with phase sensitive imaging. This was originally proposed to improve grey- to white-matter contrast in brain MRI (22), however it can also be used accentuate contrast between other tissues, like the spinal cord, where T1 times vary, to produce contrast between lesions and NASC. Image data can be reconstructed in two ways – using PSIR reconstruction or using magnitude reconstruction. The latter is most commonly used.

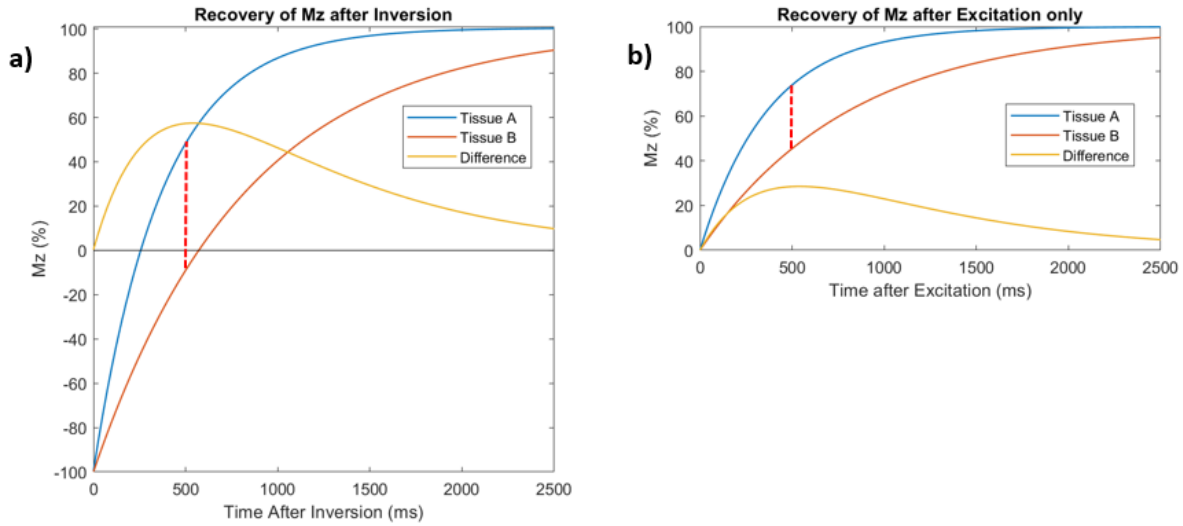

Figure 5: Shows the recovery of the longitudinal magnetization with (a) and without (b) an inversion pulse. Example a) is representative of a T1-weighted inversion recovery sequence and example b) of a more conventional spin-echo sequence. The inversion sequence produces a greater difference in longitudinal magnetization between the two tissue types.

Inversion recovery sequences make possible phase sensitive inversion recovery (PSIR), also known as true inversion recovery. It is achieved by modifying image reconstruction to use phase data which is inherent in scan data. This makes it possible to determine whether The Mz in a particular voxel is in the positive or negative z-direction at the point of excitation. Once this is done, then the grey tones of the image are used to represent signals levels, where the negative maximum is black, the positive maximum is bright and the zero value is mid-grey.

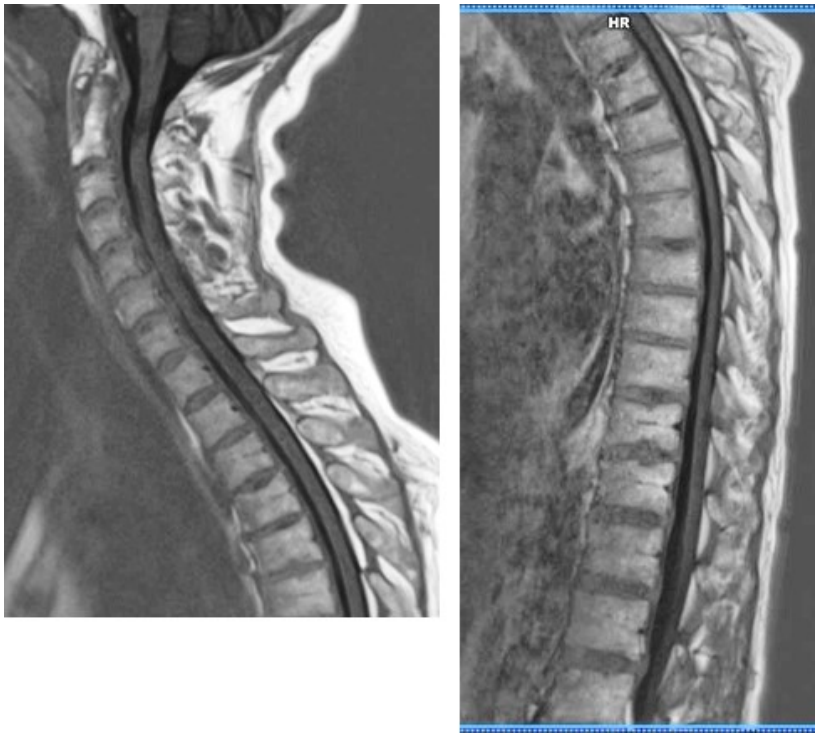

Figure 6 shows the PSIR sequence. Note the mid grey background and the vary dark CSF. A lesion seen at the C2 level appears as a dark band in mid-grey spinal cord tissue.

Comparing PSI to magnitude reconstruction, T1 contrast is particularly improved when the two tissues in question are respectively above and below the null point when the 90 degree excitation pulse is applied. In magnitude reconstruction, the difference in signal values between lesions and NASC can be very close or equal to each other, as magnetization polarity is not taken into account: no contrast will be apparent. This is shown in Figure 7a) . This does not occur with PSI reconstruction, as shown in Figure 7b).

When inversion sequences of this type are magnitude reconstructed they resemble STIR images. Inversion times are somewhat longer than with STIR, typically being around 400ms on 3 tesla imagers. Magnitude images should show more selective suppression of NASC signal. As such, this type of image can be referred to as a white-matter-suppressed (WMS) image. Spinal cord T1 times in healthy volunteers have been measured to be roughly 850-1000 ms at 3T (23). This would suggest a longer inversion time of 570-670ms using Eq. 1, to better suppress NASC. But the choice of 400ms inversion time is appropriate in order to avoid potential reduction in contrast, when PSI reconstruction is not utilised, as shown in Figure 7b). This can occur as T1 times will vary from person to person.

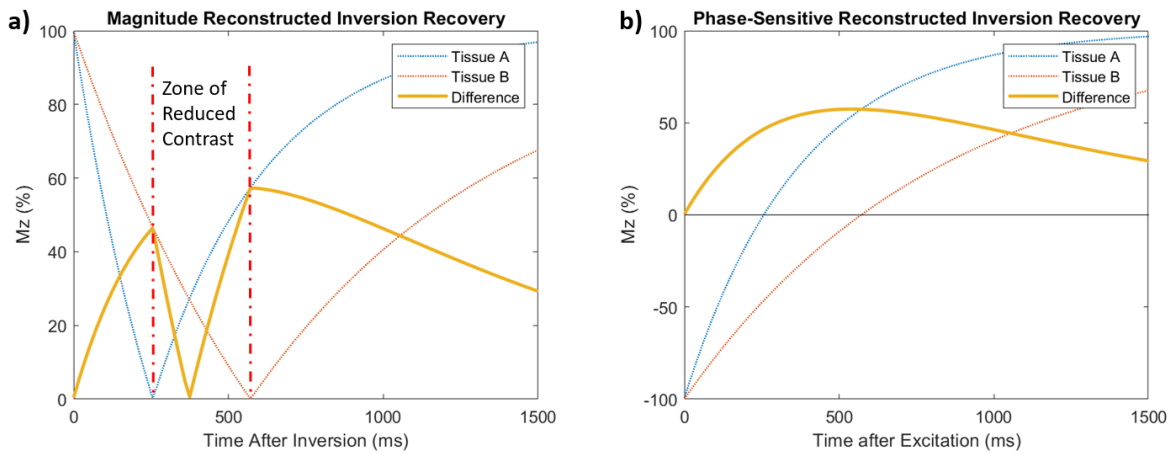

Figure 7: Shows the difference (orange line), in Mz between two tissues after inversion. The difference is equivalent to image contrast and can be seen to have a hole around 400ms (a), when magnitude imaging is employed. This does not occur with PSI reconstruction (b).

It is important to note, that when Mz is zero in a voxel prior to and hence after excitation, the voxel is given a mid-grey intensity value by the algorithm. So although there is no MRI signal present, the image contains a medium grey voxel, which can seem counter-intuitive. For this same reason, the background of PSIR images is also mid-grey. This sets this sequence and reconstruction method apart from most other MRI images.

Phase Sensitive reconstruction

Reconstruction of PSIR images requires the use of complex data from the MRI acquisition. This MRI data contains both the magnitude and phase information. Polarity is of immediate interest for reconstruction: whether the magnetization is parallel or anti-parallel to the z- or longitudinal direction. This must be determined to produce a PSIR image.

Phase errors occur begin to accumulate in the magnetization after the excitation pulse flips it into the transverse plane, complicating the determination of polarity and the subsequent reconstruction of the PSIR image. Phase errors have many causes including variations in B_0 or B_1 (RF), eddy currents, motion and flow. These errors must be accounted for in order to remove their effects and reconstruct a useful image. Initial attempts to do so were limited in success and required either several scans of varying inversion times or a reference scan. This incurred a time penalty limiting its usefulness. Later, methods for calculating polarity and estimating phase errors from real image data alone were invented (24, 25). These methods assume that phase errors change gradually going through the image, whereas polarity should change abruptly at margins between tissue types. This is utilised with a seed-growing algorithm which defines areas likely to have the same polarity. Many seed pixels are placed in the image in order to ensure complete coverage (21).

This method has been shown to produce phase sensitive images reliably and without the need for reference scans or multiple inversion scans (25).

Three-Dimensional Fast-Spin-Echo (3D-FSE) Imaging

Three dimensional imaging offers higher resolutions within a clinically acceptable timeframe. One millimetre isotropic voxel sizes are a typical level of resolution for 3D imaging. The thin slice thickness achievable with these scans, at typically 1mm thick, gives potential to detect more MS lesions. 3D isotropic imaging also allows reconstruction of image datasets in any plane, to aid in the visualisation of complex human anatomy (26).

FSE imaging is a very widespread technique in MRI, providing relatively artefact free images. It is a standard workhorse in clinical MRI. The combination of 3D imaging and FSE, is a more recent development in MRI. Improvements in certain technical aspects of the scan technique allowed manufacturers to offer 3D FSE, to offer single slab 3D imaging of sufficiently large volumes from about ten years ago.

3D-Imaging

3D imaging requires the filling of a three dimensional k-space. In comparison with 2D imaging, the number of k-space points is multiplied by a factor equal to the number of slices. Given that many thin slices may be required to encompass a sizeable volume, the number of slices can large – perhaps 64 slices. There are therefore very many k-points to be measured compared to 2D imaging.

The technique enjoys an improved signal-to-noise ratio, because a volume of tissue is excited rather than a relatively thin slice giving more overall signal. The volume, called a slab, contains contiguous slices which are spatially encoded in the slice direction by the use of a second phase encoding gradient, orthogonal to the other two imaging gradients (16). This greatly improves resolution in the slice direction. However Gibb's or truncation artefacts can now occur in the slice direction, as spatial signal variations on this axis are now determined by phase encoding, as opposed to spectral slice selection. Slice thickness must be small to mitigate this.

3D gradient echo sequences have been in use for a number of decades and is well suited to 3D imaging because the technique allow many readouts within a short space of time. It is however prone to artefacts due to static field or RF inhomogeneity. Spin echo or FSE is not prone to this, due to the use of refocusing RF pulses. For this reason, FSE in two-dimensional form continues to be the mainstay for MRI imaging. However, until about ten years ago, 3D FSE was not practical when attempting to image larger volumes in a single slab.

3D-FSE

The technical improvements and innovations which have allowed 3D-FSE to be achievable within clinically acceptable timeframes are as follows:

- Reduction of IES
- Very long echo trains.
- Storage and recall of transverse magnetization in the longitudinal direction
- Variable angle refocusing pulse schemes
- Appropriate 3D k-space filling methods

The use of long echo trains in particular has a strong effect on image contrast – possibly to the detriment of lesion contrast in spinal cord images. How these possible effects on image contrast are mitigated will be examined with in this section.

Reducing IES

3D imaging excites a slab of tissue, containing many slices. The excitation RF pulse can be either slab selective or not, when only a single slab is being imaged. When inversion pulses are used, these can also be slice selective or not. Slab selection is achieved in the same way as slice selection in 2D imaging, where a slice selection gradient is applied whilst an RF excitation pulse of appropriate bandwidth is used. Spatial variation of the Larmor frequency in the scanner means that only the magnetization in the desired slab of tissue is excited. Slab selection could also be used in the case of refocusing pulses. However, as explained above, this may require too much time, as the spatially selective refocusing RF pulse prohibitively lengthen acquisition times. In order to reduce IES, the refocusing pulse is non-spatially selective: no gradient is used in conjunction with refocusing. This

reduces IES from perhaps 8.5 milliseconds to 3.9. This reduction in turn allows for an important reduction in acquisition times.

3D-FSE can suffer problems with the phasewrap artefacts and certain free-induction-decay artefacts in the slice direction. These are largely overcome by the use of spatially selective excitation followed by an initial refocusing pulse, which is slab-selective. Subsequent refocusing pulses in this echo train are not slab selective to minimise IES.

Very Long Echo Trains and Image Contrast

In 2D-FSE imaging, echo train lengths of 15-30 are common. This allows an equal number of k-space lines to be filled within a single repetition cycle and is synonymous with turbo-factor. The acquisition time is reduced by this factor. In 3D-FSE, turbo-factors of 90 to 300 can be used. This is an order of magnitude greater and in combination with shorter IES, reduces acquisition time for 3D-FSE such that it becomes practical to use.

This does however give two problems: controlling image contrast and blurring artefacts. Under conventional circumstances, the longer the echo train, the longer the effective TE of the pulse sequence. With very long echo trains, this can lead to rather extreme T2 weighting detrimentally affecting image contrast.

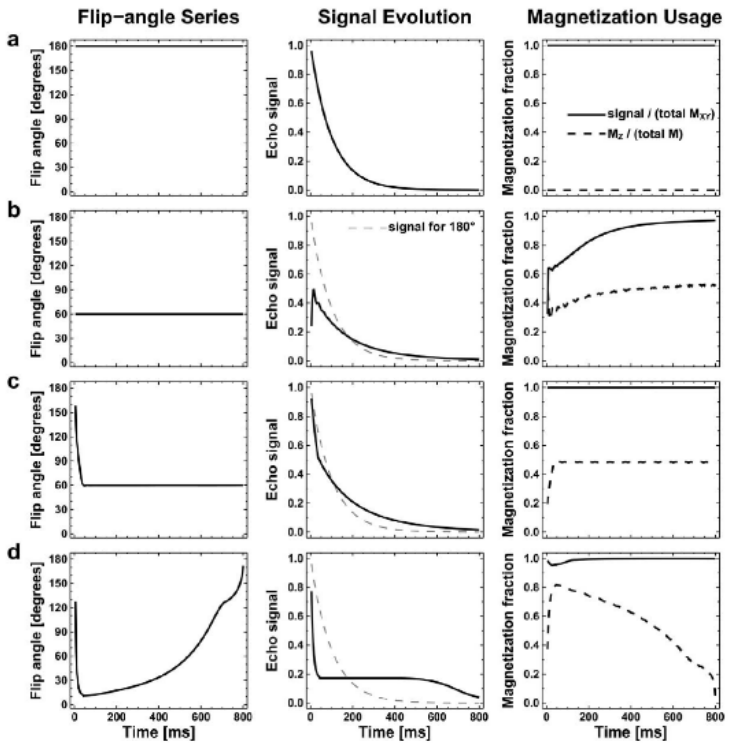

Figure 8: shows the evolution of transverse magnetization with separate flip-angle

RF-pulses are normally classified by their roles, which are usually related to their flip angles: 90 degrees for excitation, 180 degrees for refocusing. In 3D-FSE, very many RF pulses are used with an intermediate flip angle. A single pulse can have one, two or all three functions: excitation to flip magnetization into the transverse plane, refocusing to produce a spin-echo and finally storage and recall of magnetization to and from the z-direction to create stimulated echoes. The extent to which a given RF pulse fulfils these roles is dependent on its flip angle and its position in the pulse sequence. In addition, with 3D-FSE, RF flip angles can be varied during the repetition cycle in order to control the relative levels of storage, recall and refocusing to mitigate contrast and blurring problems. How much it does of each is determined by its flip angle and orientation.

Transverse magnetization produced after the initial excitation is again "stored" in the z-direction by a storage RF pulse of 90 degrees after a period of time equal to TE/2. Importantly, this magnetization decays due to T1 effects, rather than T2, thereby slowing its decay significantly. It can be "recalled" from the z-direction in full or in part by RF pulses of intermediate flip angles. After recall, transverse components will rephase over a period of time (TE/2) to produce a stimulated echo for measurement. Contributions to the signal are also made by spin-echoes, so the resultant signal is a mixture of both. Decay effects due to the long echo train can now potentially be greatly reduced, as magnetization has been stored in the z-direction and has decayed to a much lesser degree.

There are separate flip-angle schemes, which can be used to produce images of the desired contrast. These can be divided into three overall groups, shown in Table 1.

Table 1 showing flip angle schemes which can be used to obtain a particular image contrast.

Desired image contrast	**Flip-angle scheme**
Strong T2	High constant flip-angles
Short effective TE (T1 or PD contrast)	Variable flip angles Or large initial flip-angles with subsequent constant and low flip-angles
Long effective TE (T2 contrast)	Variable flip angles Or large initial flip-angles with subsequent constant and low flip-angles

K-space is filled to optimise efficient filling of all points in the three dimensional k-space. To make good use of the very long echo train, different slices are filled with data gathered in a single echo train. Radial k-space filling, starting in the centre of k-space gives images of shorter apparent TE. Linear filling produces long effective TE images, as the centre of k-space is filled relatively late. This is further accentuated by linear filling of k-space: a more conventional filling method, where spatial frequencies of a high value are filled first, moving on to low spatial frequencies and then to high spatial frequencies again.

Variable flip angles: Controlling Image contrast and blurring artefacts

As described above, the extent to which an RF pulse refocuses or recalls is dependent on its flip angle. In 3D-FSE, this is typically varied throughout the echo train in order to even out signal levels between all (stimulated and spin) echoes in the long echo train. Flip angles are varied so that magnetization stored in the z-direction is gradually recalled during the echo train to achieve this. This is shown in Figure 8, where it can be seen that in comparison with conventional constant flip angles of 180 degrees (5a), variable flip angle schemes even out signal levels between echoes during very long echo trains (5d). The magnitude of the stimulated echo measurement will still reflect that of the initial transverse magnetization after the first excitation pulse but effects due to decayed transverse magnetization will be greatly reduced.

A very important consequence of this is that the image produced has the appearance of a sequence with a very much reduced effective TE. Tissues of shorter T1 times, such as the neural tissue of the spinal cord and lesions contained within would normally give very little signal with very long effective TEs. Image contrast is transformed from highly T2-weighted to being moderately T2 weighted.

It is reasonable to assume that shorter effective TE times would also increase signal from tissues with shorter T2-times than CSF, namely MS-lesions. It is however difficult to ascertain a precise

effective TE of a given 3D-FSE sequence in order to compare with 2D-FSE images. It is generally the case that the time point after excitation where the centre of k-space is filled is given as the effective TE. With the use of store-recall pulses, the contrast the 3D-FSE image appears to have will seem to reflect an image with a rather shorter TE time than the value typically shown on the scanner. Even with careful setting of TE on the scanner, it is difficult to predict if a 3D-FSE can produce images with good lesion contrast for detecting MS lesions.

Blurring occurs because of T2 decay in the transverse magnetization. K-space lines which are filled relatively late in the echo train are measured when the magnitude of this magnetization is markedly reduced. This is typically where k-space lines from the periphery of k-space are filled, thus containing measurements of high spatial frequencies. As the signal level here is reduced, then the level of resolution in the resultant image is reduced. When variable flip angles are used, the evening out of signal level between different echoes in the long echo train mitigates blurring significantly. 3D-FSE should therefore still have the potential to visualise even small MS lesions without blurring being a problem.

3D-STIR-FSE

2D-STIR is known to be sensitive to MS-lesions. A three dimensional implementation of STIR may have potential to detect more MS lesions than 2D-STIR due to its higher resolution, particularly in the slice direction. This may allow smaller and more laterally placed lesions to be detected. CSF is bright, the spinal cord is a dark grey and fat is dark. MS lesions would appear as zones of high intensity in the spinal cord.

The sequence is however untested. CNR is hard to predict due to the unknown signal response from MS-lesions in relation to NASC when very long echo trains with variable flip angles are used. One aspect of long echo trains is that effective TEs are usually quite long – above 100ms. This could be a disadvantage as it is known that PD-weighting can visualise all pathologically confirmed lesions post-mortem (6). A PD-weighted 3D-FSE variable flip angle scheme could be applicable in order to achieve a lower effective TE. However this is not always available on implementations of the 3D-FSE sequence as provided by MRI manufacturers for spinal cord imaging. They may be more widespread for musculoskeletal imaging, particularly for knee MRI, but these implementations may not be adaptable to spinal cord imaging, as they may not be slab selective or they may not be optimised for use with spine imaging coils. Long effective TE flip angle schemes are more standard for spinal imaging.

Measuring Contrast in Diagnostic Images

It is clear that the ability to detect MS lesions in the spinal cord with MRI is a very important diagnostic tool. An important way of measuring the image quality is to assess signal differences between pathology and normal anatomy. This contrast between tissue types can be measured empirically. There are various ways of quantifying lesion contrast. What is required is a metric that gives a good indication of signal intensity differences between anatomical and pathological areas. This can be done subjectively, by asking raters to grade lesion conspicuity or more quantitatively by measuring image intensities.

Quantitative contrast metrics measured using image intensities can be calculated either with or without taking image noise into account. In practice, regions-of-interest (ROI) are defined by selecting particular areas in the image. From these, average image intensity, size and standard deviation of image pixel intensities in the region can be obtained.

Contrast-to-(spinal) cord ratios (CCR) do not include image noise as a factor but do give a reasonable idea of lesion detectability. With respect to lesions detection in the spinal cord, two formulas can be used:

$$CCR = \frac{S_L - S_{NASC}}{S_{NASC}}$$

Eq. 2 CCR1

This formula can be problematic to use when images with negative signal values, such as PSIR, are studied. The signal value of normal appearing spinal cord (NASC) can be close to zero, especially when the sequence is optimised to null the signal from healthy spinal cord. In this case CCR on PSIR can become inappropriately very large as the denominator approaches zero. An alternative equation has been utilised (27, 28):

$$CCR = \frac{|S_L - S_{NASC}|}{S_L + S_{NASC}}$$

Eq. 3 CCR2

Where S_L is the image intensity in a lesion and S_{NASC} is the signal from normal appearing spinal cord (NASC). On the face of it, this metric seems better. But it can also be problematic when used with PSIR images. If lesion signal and NASC signal are coincidentally close to equal magnitude but opposite polarity, then similarly, the denominator acquires a very small value. This makes the CCR inappropriately large, given it can occur with small magnitude values for lesions and NASC.

Taking noise into account means utilising standard deviation (S.D) measurements. Often in imaging studies the following formula can be used:

$$CNR = \frac{S_L - S_{NASC}}{S.D_{Background}}$$

Eq. 4 CNR1

However noise in many MRI images is in reality difficult to measure. Calculating noise using separate signal and noise ROIs in a single image is problematic as noise is spatially dependent, especially when parallel imaging is used. Noise measurements will vary from true SNR. Reconstruction algorithms, which can vary between images, also affect noise differently hindering comparisons. Noise is most correctly measured by repeating acquisitions, preferably many times, but this is rather cumbersome when imaging patients (29).

An alternative is to use the standard deviation in the ROIs of lesions and NASC. This method avoids some of the problems of Eq. 4 whilst still being feasible for use in diagnostic accuracy studies. One possible drawback is that heterogeneity in tissues will also contribute to S.D. This can be somewhat countered if ROIs are placed identically on the two images to be compared as the same varying tissue is measured on both. On the other hand, different weightings may lead to tissue heterogeneity affecting S.D differently, giving one image weighting an unfair advantage. However the following formula can perhaps be considered the least worst option:

$$CNR = \frac{Signal_{Lesion} - Signal_{NASC}}{\sqrt{S.D^2_{Lesion} + S.D^2_{NASC}}}$$

Eq. 5 CNR2

The above formula has been used in several imaging studies including of the spinal cord and PSIR images (27, 30, 31). It is noted that the formula converts ROI S.Ds to variance before adding them and converting back to S.D with the same units as signal.

Imaging Research: MS spinal cord lesion detection

A peer-reviewed literature search for diagnostic accuracy studies using pubmed was carried out using combinations of the search terms MS or Multiple Sclerosis, MRI, Spinal or neck or cervical, Cord, sequence, lesion or lesions. Sixteen articles were identified from the past eleven years (32-47).

Three studies scanned both the cervical and thoracic spinal cord. The remaining studies scanned the cervical spinal cord only. Three studies used 1.5T magnets exclusively. 3T magnets were used in the majority of studies – twelve, with a single study being carried out on a combination. This seems to indicate that 1.5T is a relatively under-researched area for MS lesion detection in the spinal cord, although differences between 1.5T and 3T is not the focus of this paper. Only four studies examined the use of three-dimensional sequences which have a greater overall spatial resolution.

STIR Studies

Since 2007, the relative diagnostic usefulness of STIR, as compared to other sequences has been investigated seven times. In five studies, STIR was found to detect the highest number of lesions (34, 39, 41, 46, 47). In the two studies with a reference standard, STIR had a specificity lower than that of another included sequence (36, 39). T1-weighted 3D MPRAGE scanning and white-matter suppressed T1 imaging (a variant of PSIR) have been found to reveal more lesions than STIR, each in a single study (36, 43). Interrater agreement was evaluated in four of the studies, where STIR was superior in three, but inferior to a form of STIR on the fourth. Subjective evaluation of artefacts were not directly evaluated in any study with STIR. A study of three-dimensional STIR was not found in the literature search.

T2 Studies

Many MS lesion detection studies in the spinal cord compare one or more test sequences with T2, sometimes with T2 in combination with other sequence types. T2 was included in twelve studies where it was compared to another image weighting. In every case it was found that the other sequence performed better in detecting lesions (32, 33, 37-47). A reason for this could be a generally poor lesion contrast. Seven of the twelve studies included contrast measurements and T2 was inferior in every case. Despite this T2 sequences have long been a standard sequence for spinal cord imaging, due to its generally good SNR and low prevalence of imaging artefacts. Importantly, of the four studies where specificity was estimated, T2 was found to be the most specific in three (39, 42, 44). T2 also performs reasonably well in terms of interrater agreement.

T2 weighted images were used exclusively in the only two studies which examined full-coverage axial scanning. Here it was found that axial image evaluation of the entire examined area revealed many more lesions in comparison to sagittal T2 images (35, 37).

PSIR Studies

PSIR images have been tested several times on 3 Tesla scanners (33, 39, 46, 47) and once at 1.5T (34). Poonawalla et al (47) compared PSIR with dual echo imaging and T2 in the cervical spinal cord. PSIR and STIR were rather equal in detecting lesions, but both were superior to T2. PSIR produced a better CCR.

Philpott et al (46) compared contrast metrics between PSIR to STIR and T2, also of the cervical spinal cord. Lesion detection was not evaluated. The PSIR sequence was a rather lengthy six minutes

compared to the 2 and 3 minutes of the other sequences, possibly giving PSIR an advantage. The authors attempted to compensate for this by dividing contrast metrics by the square root of the acquisition time.

Alcaide-Leon et al (39) studied PSIR over the whole spinal cord. PSIR detected more lesions in the cervical spinal cord, being superior to STIR, PD and T2 imaging. However T2 was most specific on cervical scans. In the thoracic spinal cord, STIR detected the highest number of lesions whilst PSIR detected the fewest but was most specific. This indicates differences in diagnostic accuracy for sequences, depending on whether they are scanned in the cervical or thoracic areas. Overall, interrater agreement was best for STIR and worst for PSIR, possibly indicating difficulties in utilising an unfamiliar sequence or perhaps differences in SNR or artefacts between sequences. Contrast metrics and subjective evaluations of artefacts were not reported.

Fechner et al conducted a study with 3D PSIR (gradient echo) against an image set of T2 and T1 images. PSIR detected more lesions and produced a higher CNR. No consensus reference standard was used. Interrater agreements were similar for both image sets.

However, Shayganfar et al recently conducted a study at 1.5T (34), which found PSIR to be inferior to STIR in terms of lesion detection. It is interesting to note that the echo time used on 1.5T was rather high at 80 milliseconds, which may have reduced SNR. For interrater agreement, PSIR performed reasonably compared to STIR. Contrast metrics were unfortunately not measured. This study was published after image acquisition had been completed for this paper's project.

Three-Dimensional Imaging

3D imaging has been investigated several times in recent years for use in MS lesion detection. 3D imaging can yield increased overall resolution, especially so in the slice direction (16). 3D double inversion recovery (DIR) (40), T1-MPRAGE (43) and 3D-PSIR (33) have been evaluated. One of these is the study by Fechner discussed above.

3D-DIR was superior to T2 sagittal and axial images in terms of number of lesions and detected contrast metric, but suffered more from low image quality, due to motion artefacts from swallowing. Three scans were discarded from the study due to other artefacts including aliasing and B1 inhomogeneity artefacts due to patient obesity.

T1-MPRAGE outperformed two-dimensional STIR and T2 in terms of lesions detection and CCR. In this study however, all sequences were reviewed together and at the same time. It could therefore be the case, that some sequences have had their sensitivity increased by being compared with other sequence types: a lesion seen on one sequence would obviously lead the eye when looking at another.

3D imaging is a lesser explored avenue in MS lesion detection and has not been explored on 1.5T scanners at all. The technique may confer certain advantages due to its higher overall resolution independently of image weighting. In particular, STIR has not been investigated in a 3D implementation.

Increased resolution reducing partial artefacts can detect more lesions

Laterally located and small lesions in the spinal cord are especially prone to being missed on sagittal scans (35, 37). 2D sagittal imaging, often with a slice thickness of 3mm is prone to problems due primarily to partial volume artefacts. This artefact occurs when more than one tissue type is present within a single voxel. In the case of sagittal MRI imaging, the two slices positioned on the lateral edges of the spinal cord can easily be half in and half out of the solid tissue, the other half or so

being CSF. In a T2 or STIR slice of normal thickness, e.g 3mm, a lesion which would otherwise be of intermediate intensity but only fills half the slice thickness, would be drowned out by the CSF occupying the other half of the voxel. Even when the whole slice is within the spinal cord, smaller lesions can be missed, due to the same artefact. This is despite the fact that 2D slices can have a sub-millimetre in-plane resolution: low resolution in the slice direction is problematic.

Two studies, each with over 100 test subjects, have reported the relative visibility of lesions on sagittal scans as compared to complete axial coverage of the spinal cord (35, 37). Both found that most lesions missed on sagittal scans but seen on axial were located laterally in the spinal cord and were smaller in size. Breckwoldt et al found that 2.6 times as many lateral lesions were seen on axial images contra sagittal. The average size of additional lesions found on axial images ($7mm^2$), including those more centrally located was significantly smaller than those lesions seen on both sagittal and axial images ($16mm^2$). Galler et al found that sagittal scans detected only 57% of lesions seen on axial images. Only 46% of all detected smaller lesions of less than 3mm in diameter were visible on sagittal images, contra 93% of lesions greater than 5mm.

These results indicate that higher resolution, especially in the slice direction for sagittal images, may improve the detection rates of both smaller and laterally located MS lesions.

Research Question

A diagnostic accuracy study was carried out to investigate the relative diagnostic usefulness of several sequence weightings. Diagnostic accuracy was investigated by comparing the number of lesions detected, CNR measurements, and prevalence of artefacts and noise for each weighting. The entire spinal cord was fully covered by the use of two sagittal scans for each weighting, designated as cervical and thoracic coverage areas. Weightings were primarily compared by coverage area: cervical to cervical and thoracic to thoracic.

There are various sequences which can be utilised in the detection of MS lesions in the spinal cord:

- Sagittal 2D-STIR
- Sagittal T2
- Sagittal 3D-STIR
- Sagittal PSIR

Of these sequences, which has the best diagnostic accuracy, as measured by sensitivity, positive predictive value and interrater agreement? Additionally, which sequence shows is least affected by artefacts and noise?

Null-Hypotheses

- There is no significant difference in the number of detected multiple sclerosis lesions in the spinal cord between the four MRI sequence types by coverage area.
- There is no significant difference in the contrast-to-noise ratio of detected multiple sclerosis lesions in the spinal cord between the four MRI sequence types by coverage area.
- There are no significant differences in levels of artefacts between sequence types by coverage area.

Method

Choice of Sequence and Optimisation

Two sequences of each included image type (weighting) were scanned on every patient, one in the cervical anatomical area including four to five thoracic vertebrae and one in the thoracic area including two lumbal vertebrae.

All sequences were based on the department's MS spinal cord protocol or manufacturer's factory settings for cervical spinal cord scanning with important adjustments to slice thickness, resolution etc. These were copied and adapted for thoracic scanning. A summary of sequence parameters is shown in Table 2. A full list of all parameters is available as an appendix. All scans were performed on the same scanner (1.5 Tesla Avanto Fit, Siemens, Erlangen, Germany). All cervical sequences incorporated the use of the anterior part of the 20-channel head/neck coil. A 32-channel spine coil was also part of the setup.

A small pilot study was carried out in order to optimise the sequences in terms of resolution, contrast and signal. This was done subjectively. In order to make sequences more comparable, an effort was made to optimise scan times to approximately four minutes. Fields of view (FOV) were identical for all sequences in the same anatomical area in order to ensure identical coverage in the superior/inferior direction of the spinal cord. Thoracic FOV was markedly larger than cervical. Differences in voxel size between cervical and thoracic images were minimised by increasing in-plane resolution on the thoracic sequence. Coverage in the slice direction was identical for all sequences except for 3D-STIR, which requires more slices to ensure effective sampling. Slice

thickness for T2 and 2D-STIR was reduced to 2.5mm to improve resolution, but not PSIR due to signal constraints. In-plane resolution for T2 scans was increased markedly as there was ample signal. In-plane resolution was at least 1x1mm for all sequences.

Table 2: shows a summary of imaging parameters for all sequences.

	2D-STIR		T2		3D-STIR		PSIR	
	Cerv	**Thor**	**Cerv**	**Thor**	**Cerv**	**Thor**	**Cerv**	**Thor**
TR (ms)	3800	3800	3780	3600	3500	3500	1870	1870
TE (ms)	37	36	99	94	182	182	15	15
TI (ms)	160	160	0	0	160	160	347	347
No. Slices	11	11	11	11	64	64	11	11
FOV (mm)	256	320	256	320	256	320	256	320
% Phase FOV	84,4	100	84,4	100	100	100	100	100
Matrix	320	384	448	512	256	320	320	384
Phase resolution	100	100	100	100	90	85	85	85
Slice thickness/ Gap	2,5/0,25	2,5/0,25	2,5/0,25	2,5/0,25	1/0	1/0	3/0,3	3/0,3
Turbo Factor	19	15	30	15	170	170	4	4
No. Averages	4	3	5	4	1,4	1,4	2	2
GRAPPA Factor	2	2	2	2	3	3	2	2
Oversampling (%)	100	100	75	85	95/12,5	95/12,5	95	66
Flow Compensation	No	No	Read	Read	No	No	Read	Read
Phase Direction	A/P	H/F	A/P	H/F	H/F	H/F	H/F	H/F
Acquisition Time	3'53"	3'53"	3'52"	3'56"	3'49"	4'17"	4'14"	4'18"

Study Participants

Patient recruitment

A convenience sampling method was used. In order to maximise the number of lesions in the patient cohort, the following was done: the departmental Roentgen Information System (RIS) was used to search for MRI examinations of the spinal cord. First, the exam report was read to ensure that lesions had been found. Next, the patients' electronic journal was used to check that the diagnosis of MS and/or sub-type had been given. The RIS system was checked backwards in time from the date of the search, so that patients who had more recently been scanned would be found first. This process was repeated several times roughly at two month intervals. Invitations were subsequently sent out to potential participants. Only those participants who replied to the invitation were screened for suitability in terms of the inclusion/exclusion criteria. Those found suitable were enrolled in the study. Data from electronic journals pertaining to MS type and disease duration were collected. Examination descriptions were not produced.

The project protocol was approved by the science ethics committee (Videnskabsetisk Komité) for the Zealand region of Denmark. There was full compliance with the Helsinki Declaration on ethical

principles for medical research involving human subjects. All patients gave informed written permission to take part, including the use of electronic journal information.

Inclusion Criteria

- MS diagnosis
- Known to have spinal cord MS-lesions
- Able to stand
- Over 18 years of age

Exclusion Criteria

- MRI conditional or MRI unsafe implants.
- Pregnancy

Scanning of Patients

Time was taken to ensure patients were comfortable on the scanner table. All patients were reminded of the importance of lying still. Sequences were scanned in a predetermined random order to reduce possible systematic bias. Automatic spine labelling was not used. All sequences within a patient were scanned with precisely the same orientation and positioning in the cervical and thoracic areas respectively. The cervical and thoracic scans for each weighting were scanned concurrently. The order of image weightings was randomly chosen for each patient, with a reasonable distribution of possible combinations.

Diagnostic evaluation

Two qualified radiologists took part in the evaluation. Rater A was a departmental radiologist with half a year's experience as a specialist in neuroradiology. Rater B was a consultant level neuro-radiologist with four years experience at this level. Raters were blinded to all clinical data. All images were evaluated on diagnostic displays.

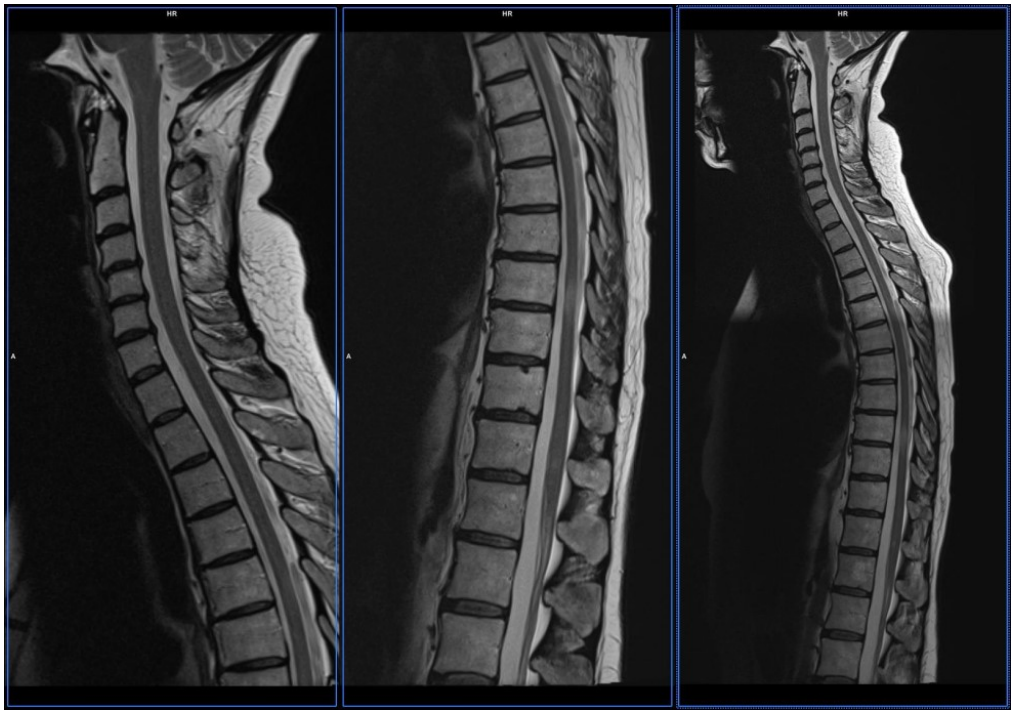

Figure 9 shows an example of the three images presented to raters at initial evaluation for T2 weighting. These are the cervical and thoracic area coverage images and a reconstructed composed image.

All examinations were pseudonymised by assigning a patient ID producing a dataset for all patients with all sequences under each patient. All sequences were subsequently divided into four further groups by weighting. For each weighting each patient was assigned a new randomised patient ID and this was used as a proxy to determine the order in which patients would be presented under diagnostic image evaluation. Order of presentation was different and random for each weighting. No patient had the same ID in any of the five groups. Reconstruction of 3D-STIR was not done at this stage.

Three sets of images were presented for each patient and each weighting: a cervical and thoracic sequence and also a composed image set as shown in Figure 9. This is a full-length image of the spinal cord automatically generated by the scanner by combining the cervical and thoracic image sets. Evaluation forms were produced for the study with patient ID, level in the spinal cord and artefact evaluations clearly marked.

Participating radiologists were informed via written documentation as to the purpose of the study and the various characteristics of the included sequence types. This included descriptions of image weighting, example images and the expected signal intensity on PSIR. Radiologists were also instructed that a meaningful amount of time should pass between readings of different image groups. Explanations were given pertaining to the grading of artefacts and image noise. Individual evaluations were conducted only once for each group.

Consensus Review

As a substitute for a golden standard, a consensus review was conducted. A lengthy period after individual rater evaluations, both participating radiologists met to conduct a consensus review of all patients with the addition of an arbitrator. All sequences were used and axial reconstruction of the 3D-STIR was also included and utilised in this session. These reconstructions were used help characterise potential findings on the sagittal images, for example, to help establish whether a high signal intensity area was within the spinal cord or not. An example of axial reconstruction is shown in Figure 10.

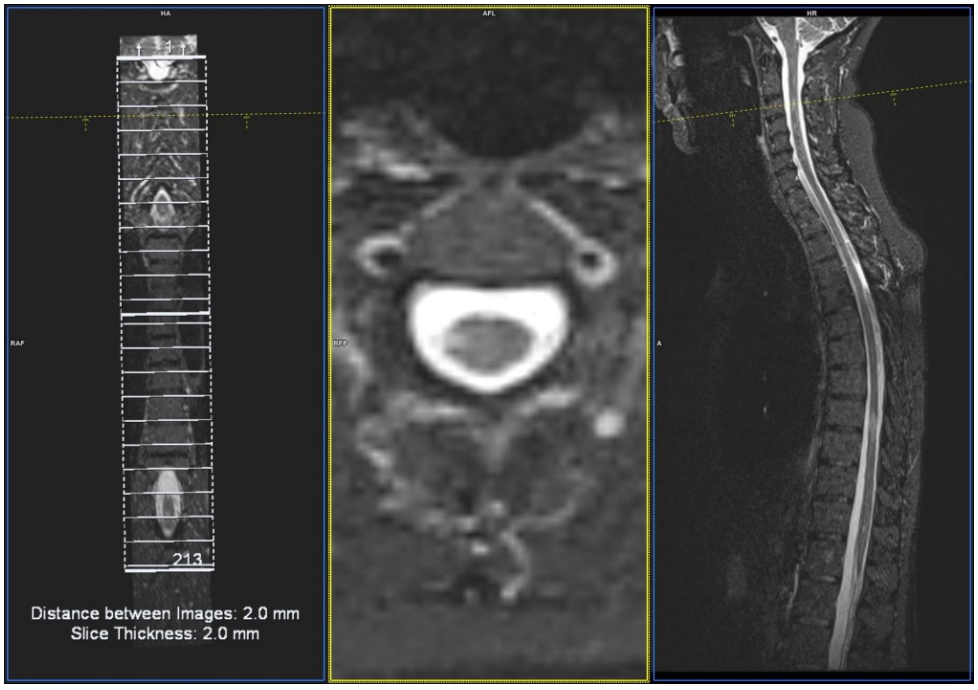

Figure 10: shows an example of axial 3D reconstruction of 3D-STIR. A small lateral lesions can be seen on the axial image. This lesion was also readily visible on sagittal images.

Adjustment of Lesion location

Upon review of the results from the lesion detection part of the study, it became clear that some variations in the assignments of anatomical locations were due to random inter- and intrarater variation. As the study focuses on the detection of lesions and not their position per se, locations were adjusted after the fact, where it was clear that the same lesion had been seen on two or more sequences, but had been given a different location due to rater variation. This was to remove random variation as a factor from sensitivity, specificity and agreement calculations. An example of the way lesion locations were adjusted is shown in Figure 11.

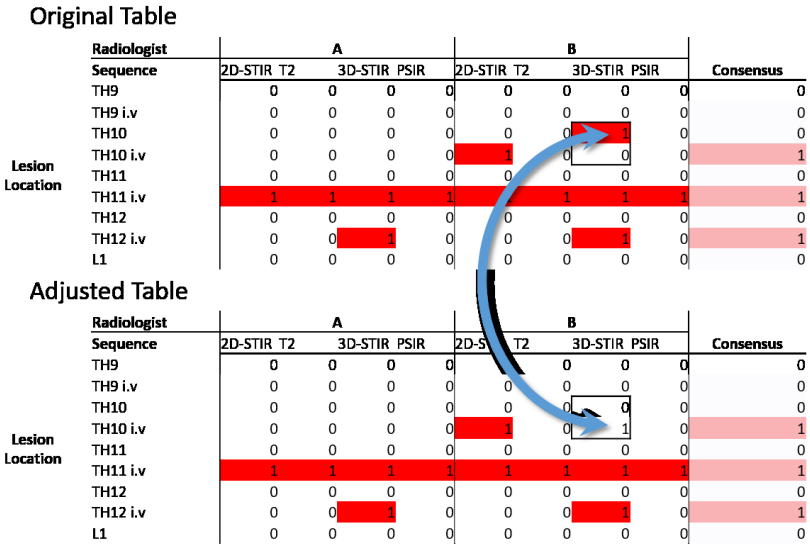

Figure 11: A screenshot of two versions of lesion detection data showing how the locations of lesions were adjusted. In the upper table, a lesion was noted at the 10th thoracic level on 3D-STIR and on the 10th intervertebral level on 2D-STIR by rater B. Consensus review placed a lesions at the 10th intervertebral level. As no other lesions are apparent in the local area, it is logical to assume that the 3D-STIR lesion is the same as the one seen on 2D and consensus at the 10th intervertebral level. The adjustment has been carried out on the lower table.

Contrast-to-Noise Measurement

CNR was calculated using Eq. 5. Due to overlap between sequences, some lesions between the 3rd and 6th thoracic levels were visible on both cervical and thoracic sequences. It was decided arbitrarily that for all patients, lesions would be evaluated on cervical sequences down to the 4th thoracic intervertebral level and more inferiorly located lesions on thoracic sequences.

Lesion locations as determined by consensus review were used to identify lesions for CNR measurement. Each lesion was reviewed on all sequences in order to ascertain which weighting showed the lesion to its greatest extent. This image was used to carefully define the lesion ROI. All ROIs were copied and pasted onto the other three sequences, with identical shape and size. Adjustments were made for possible positioning mismatches due to gross patient movement or other effects. This was carefully done by considering lesion positions in relation to anatomical markers such as the corner of a nearby vertebral body. ROI mean intensity, standard deviation and size was recorded for all lesions.

NASC signal intensities were also measured, by the placement of a ROI immediately adjacent to each lesion. Each lesion was matched too an NASC ROI in this manner. Each NASC ROI had an identical

placement on all sequences, in the same manner as lesion ROIs. An image example of ROI measurements is shown in Figure 12.

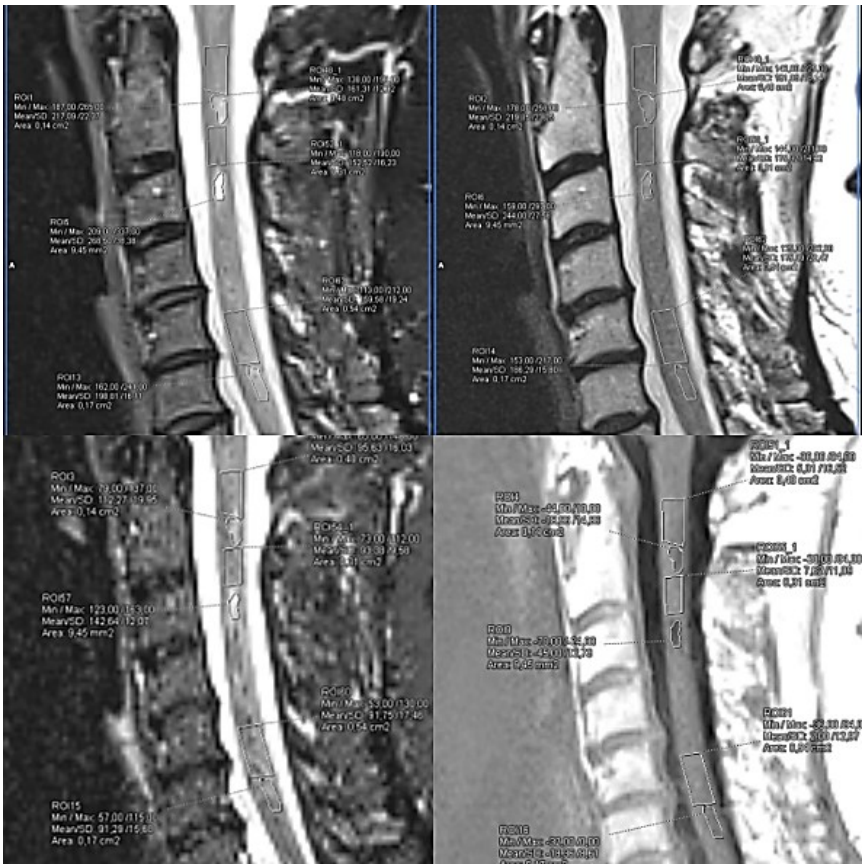

Figure 12: shows the placement of lesion and NASC ROIs on cervical scans from a single patient.

Subjective Evaluation of artefacts and Noise

Flow and motion artefacts and noise were subjectively evaluated on a four point scale from zero to three using the following grades as shown in Table 3. This was done in conjunction with individual rater lesion detection, so that each weighting was evaluated separately and were similarly randomised. No separate evaluation of cervical and thoracic images was carried out. Therefore, each patient's image set, consisting of cervical, thoracic and composed images was evaluated as a whole.

Table 3: Grading Criteria used to evaluate artefacts and Noise.

Grade	Criteria
0	Artefact / Noise not apparent
1	Artefact / Noise does not affect diagnosis
2	Artefact / Noise affects diagnosis somewhat
3	Artefact / Noise meaningfully affects diagnosis

Data Analysis

For lesion detection, sensitivity and positive predictive value (PPV) were calculated to compare the performance of individual image groups to each other and the consensus review. The Cohen's kappa ratio was used to evaluate interrater agreement. The McNemar test was used to test for significant difference in lesion detection performance between sequences.

For evaluation of artefacts and noise, within rater comparisons were made between pairs of image weightings. Simple counts of number of image sets of a certain grade were collated. The Sign test f was used to test for significant difference as this can be used for ordinal data.

For CNR measurements, it was surmised there were large variations in the number of lesions per patient. A linear mixed models approach was used to calculate adjusted mean CNR values for each sequence and to test for significant differences between sequences. This was to reduce the overall effect that the few patients with very many lesions would otherwise have on the average CNR values for the cohort. The sequence type was modelled as a fixed effect and patient ID modelled as a random effect. In order to ensure easier comparison between PSIR and the other sequences, PSIR-CNR values were multiplied by -1, to account for the fact that lesions are darker than background tissue with this weighting.

Results

Twenty-five patients participated in the study. MS type age and number of years since diagnosis are shown in Table 4. All results tables and figures are also attached as a separate appendix.

Table 4 showing data on study participants

MS Type	**Number Participants**	**Female/ Male**	**Average Age (Yrs)**	**s.d**	**Average time from diagnosis (Yrs)**	**s.d (yrs)**
CIS	3	3 / 0	54.7	3.21	2.00	5.859465
PPMS	1	1 / 0	65	-	3.00	-
RRMS	18	13 / 5	49	9.13	8.53	7.210798
SPMS	3	3 / 0	53.3	12.90	19.67	7.653975

All scans were successfully obtained with no repeat scans. For all patients the cervical images covered from above foramen magnum to beyond the 4th thoracic intervertebral level. Thoracic images covered from the 3rd thoracic vertebral body to and including the 2nd lumbal vertebra. All image sets were evaluated by both raters.

Analyses of lesions, sensitivity, specificity and CNR were done separately for the cervical and thoracic areas of coverage. These are separated arbitrarily with the dividing point being between the Th4 intervertebral level and Th5. A number of lesions were visible on both cervical and thoracic scans due to the necessary overlap so this division was done to simplify analysis. However it should be made clear therefore, that a lesion at the Th3 level was always evaluated on what is referred to as a cervical scan. Likewise, lesions referred to as thoracic or found on thoracic scans in the analyses cannot be above the Th5 level. The division is shown in Figure 13, where the number and location of lesions detected at consensus by MS type is also shown.

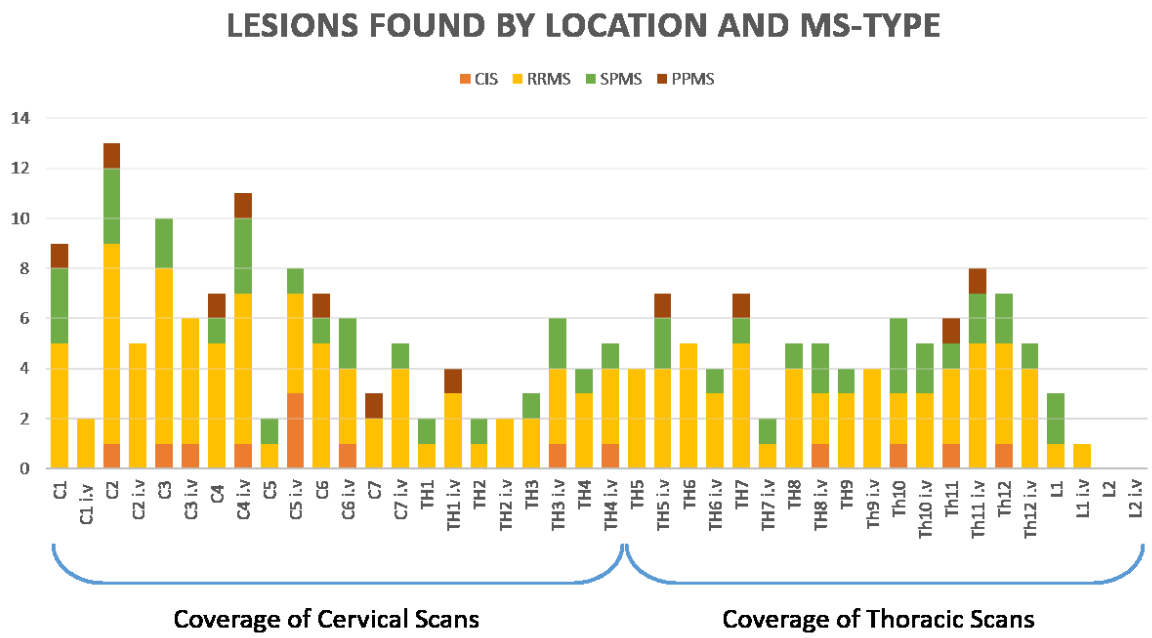

Figure 13 shows the number of lesions detected at consensus by location and MS type. The diagram also indicates which sequence a lesion found at a particular level was evaluated on for detection and CNR measurements.

Lesion Detection

At consensus review using all sequences 210 lesions were found: 122 lesions were found on cervical scans from the level of C1 to Th4. 88 were found on thoracic scans from Th5 to L1. The greatest number of lesions were detected on the 2D-STIR weightings for both raters on both anatomical (cervical and thoracic) scans. Rater A detected more lesions on 3D-STIR than T2. The opposite was the case for rater B. The PSIR weighting detected fewest lesions for both raters and area coverages. These results include false positives, i.e lesions not confirmed at consensus review and are shown in Figure 14.

Relatively many lesions were found in the SPMS patients in the study. RRMS patients made up three-quarters of the study group and had 133 lesions in total. Only one PPMS patient participated. This information is summarised in Table 5.

Table 5 Number of lesions found for each MS subtype by consensus.

MS Type	**Number Participants**	**Total number lesions)**	**Average number Lesions (S.D)**
CIS	3	14	4.7 (2.5)
PPMS	1	11	11
RRMS	18	138	7.7 (4.8)
SPMS	3	47	15.7 (2.3)

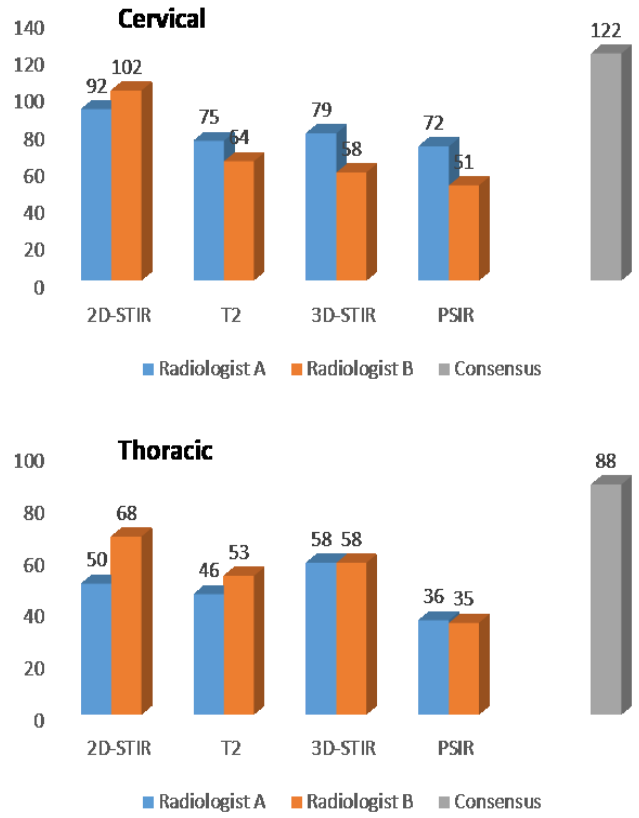

Figure 14: a bar chart of number of lesions detected by anatomical area per sequence and by consensus using all sequences.

Sensitivity and PPV values are shown in full in Table 6. The sensitivity was highest for 2D-STIR. Differences in sensitivity were most pronounced on thoracic scans for rater B with a sensitivity of over 70% for 2D-STIR and less than 40% for PSIR. Sensitivities varied less for rater A, however 2D-STIR still had the highest sensitivity on both thoracic (0.534) and especially cervical (0.680) scans.

Table 6: shows a summary of Sensitivities and PPVS for all sequences.

		Radiologist A				Radiologist B			
		2D-STIR	**T2**	**3D-STIR**	**PSIR**	**2D-STIR**	**T2**	**3D-STIR**	**PSIR**
Cervical	**Sensitivity**	68.0%	55.7%	47.5%	53.3%	73.8%	48.4%	42.6%	37.7%
	PPV	0.90	0.91	0.73	0.90	0.88	0.92	0.90	0.90
Thoracic	**Sensitivity**	53.4%	50.0%	47.7%	39.8%	71.6%	54.5%	50.0%	37.5%
	PPV	0.94	0.96	0.72	0.97	0.93	0.91	0.76	0.94

The detection data were tested for significant difference in the number of true lesions found between sequences. This was only done for within raters. P-values of less than 0.05 were considered significant. For cervical sequences, rater A found significantly more lesions on 2D-STIR in comparison with all other sequences ($p<0.005$). Rater B had the same result with the addition of significant difference between T2 and PSIR ($p=0.049$), with PSIR detecting fewer lesions. For thoracic sequences, no significant differences were found between numbers of true lesions detected by rater A. Rater B on the other hand, detected significantly more lesions on 2D in comparison with all other sequences ($p<0.005$). No other significant differences were found.

PPVs were fairly uniform between sequences with the exception of 3D-STIR. PPV for this sequence was markedly reduced to roughly 0.7 for both raters and areas of scan coverage. This indicates an

increased prevalence of false positives with the exception of rater B performance with the sequence in the cervical area.

Lesion size was not assessed independently on each sequence. The given sizes are therefore a compromise based on the appearance of a given lesion on all sequences. Comparisons on the sizes of lesions between sequences are therefore not possible. Instead, area of ROIs for lesions were used as a substitute. As ROIs were defined on sagittal images, the measured size reflects lesion dimensions in the superior/inferior and anterior/posterior directions.

Agreement between raters was similar for all sequences with the exception of 3D-STIR. For 2D-STIR, PSIR and T2, agreement was substantial, but only moderate for 3D-STIR using recommended descriptives (48). Agreement was greatest for T2 in the cervical area (κ=0.778 s.e=0.041). Agreement was lowest for 3D-STIR in the cervical area (κ=0.46 s.e=0.056). All results are shown in Table 7.

Table 7 showing kappa values for agreement between radiologists by sequence and area coverage.

	Cervical				Thoracic			
	2D-STIR	**T2**	**3D-STIR**	**PSIR**	**2D-STIR**	**T2**	**3D-STIR**	**PSIR**
Kappa	0.750	0.778	0.460	0.699	0.737	0.754	0.484	0.682
Std Error	0.038	0.041	0.056	0.049	0.030	0.033	0.041	0.040

Subjective Evaluation of Noise and Artefacts

The prevalence of artefacts interfering with diagnosis was generally low. A summation of evaluations of artefact and noise can be seen in Table 8.

Table 8 shows data from flow artefact evaluation.

		Motion				Flow				Noise				Inhomo-geneity
	Grade	**2D**	**T2**	**3D**	**PS**	**2D**	**T2**	**3D**	**PS**	**2D**	**T2**	**3D**	**PS**	**3D-STIR**
Radiologist A	**0**	25	24	25	23	11	12	14	25	25	25	10	25	X
	1	0	0	0	1	14	12	11	0	0	0	6	0	X
	2	0	1	0	1	0	1	0	0	0	0	6	0	X
	3	0	0	0	0	0	0	0	0	0	0	3	0	X
Radiologist B	**0**	8	19	10	12	0	2	0	0	0	0	0	0	11
	1	16	5	15	11	16	20	9	18	22	21	0	20	12
	2	0	1	0	2	9	3	13	7	3	4	16	4	2
	3	0	0	0	0	0	0	3	0	0	0	9	1	0

Noise

Rater A judged noise to meaningfully affect diagnosis for three patients' scans and somewhat affected for six patients on 3D-STIR. Noise interfered with diagnosis somewhat for six patients. No other weightings were considered meaningfully affected by noise and only somewhat affected. Rater B judged noise to be problematic for all 3D-STIR images. Noise meaningfully affected diagnosis for nine patients' scans and somewhat affected all other patients' 3D-STIR images. For the other weightings, noise affected diagnosis somewhat for only 3 or 4 patients for each weighting. On PSIR, noise affected one image set meaningfully. 2D-STIR and T2 were never considered meaningfully affected by noise.

All differences in noise evaluations between 3D-STIR and all other weightings were statistically significant for both raters. No other significant differences were found.

Motion Artefacts

Raters did not consider motion artefacts as meaningfully affecting diagnosis on any weighting.

For rater A, only 2 image sets were judged to be somewhat affected. This was for different patients; one on T2 and the other on PSIR. Rater B judged similarly, with diagnosis on two image sets on PSIR being somewhat affected by motion. T2 was judged somewhat affected for one patient.

Flow Artefacts

The raters were in broad agreement that the PSIR weighting had the lowest prevalence of flow artefacts. Rater A reported zero flow artefacts for the PSIR weighting, but visible on all other weightings. Therefore the PSIR weighting was found to have significantly fewer flow artefacts in comparison with all other weightings for this rater ($p<0.05$) using the sign test. However, flow artefacts were not judged meaningfully disruptive of diagnosis on any weighting, and only disrupted somewhat for one patient on a T2, without becoming meaningful.

Rater B judged the 3D-STIR to have the greatest prevalence of flow artefacts. The difference in prevalence between 3D-STIR and all other weightings was statistically significant ($p<0.05$). Rater B considered that flow artefacts disturbed diagnosis somewhat for 13 patients and meaningfully for 3 patients. No other significant differences were found.

3D-STIR Artefact

As scanning of patients progressed, it became clear there was some sort of inhomogeneity artefact affecting image quality for some patients on 3D-STIR. The artefact only occurred on this sequence and was included for evaluation on all sequences, so only results for 3D-STIR were collected and are shown here. For rater B, the artefact never meaningfully affected diagnosis but did somewhat affect it for two patients. Rater A did not evaluate the artefact. An example is shown in Figure 15. Here it can be seen that the overlap between cervical and thoracic images goes some way to alleviating the artefact on the composed image.

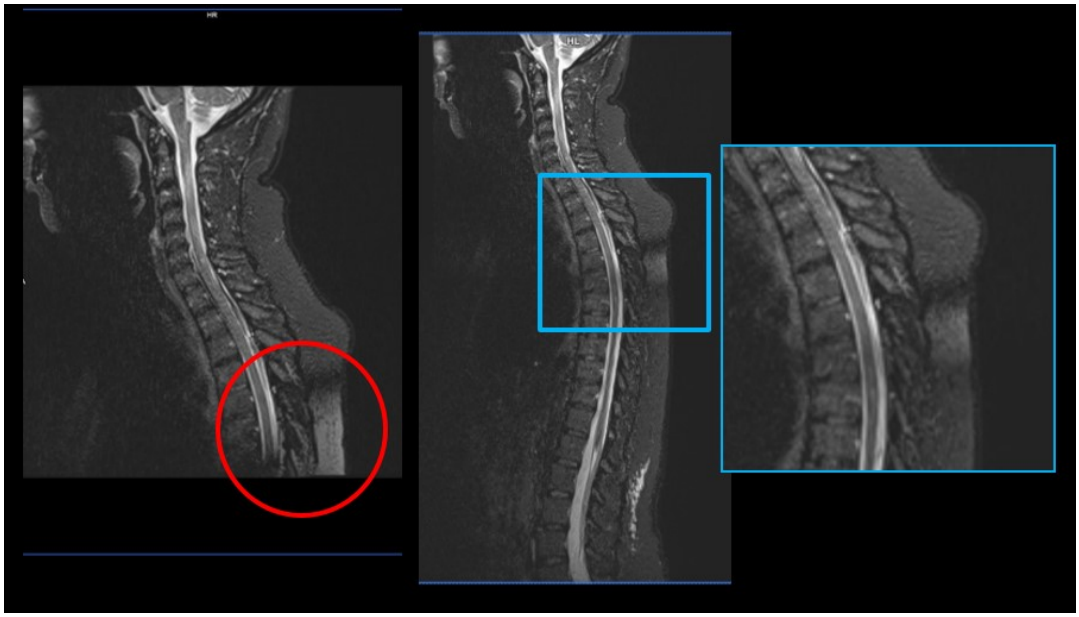

Figure 15: an example of the inhomogeneity artefact seen on 3D-STIR, outlined with a red circle. On the composed image (centre and right) the overlap between cervical and thoracic images compensates for the artefact and also the reduced signal apparent towards the edge of the field of view.

CNR ROI Measurements

Differences in calculated adjusted means for CNR are shown for all sequences in Table 9. Boxplots for CNRs are shown in Figure 16. The highest mean was on 2D-STIR and the lowest on T2 for lesions and on both cervical and thoracic images. 2D-STIR CNR was significantly greater compared with all other sequences. T2 was significantly worse than all other sequences apart from the thoracic PSIR. There were only small, non-significant differences between 3D-STIR and PSIR sequences.

Differences in CNR Adjusted Means

Cervical Scans					
		2D-STIR	*3D-STIR*	*PSIR*	*T2*
	Adjusted Mean	1.798	1.154	1.168	0.926
Comparison Sequence	*2D-STIR*		0.643 *	0.630 *	0.871 *
	3D-STIR	-0.643 *		-0.014	0.228 *
	PSIR	-0.630 *	0.014		0.242 *
Thoracic Sequences					
		2D-STIR	*3D-STIR*	*PSIR*	*T2*
	Adjusted Mean	1.696	1.161	0.978	0.951
Comparison Sequence	*2D-STIR*		0.534 *	0.718 *	0.745 *
	3D-STIR	-0.534 *		0.184	0.210 *
	PSIR	-0.718 *	-0.184		0.027

*Table 9 Table of Differences in CNR between Sequences. Means for individual sequences are shown in red. Significant differences are marked *. Level of significance was p-value < 0.02.*

There are a number of outliers and measurements where CNR values are less than zero, indicating instances where NASC and lesions had an opposite to expected values: NASC brighter than lesions, except on PSIR, where NASC had a lower signal intensity in some cases. This is a counterintuitive

result, probably due to experimental error. There are 17 of these instances for all sequences out of a total of 840 measurements. As such this was considered a minor effect, so no effort was made to correct these instances.

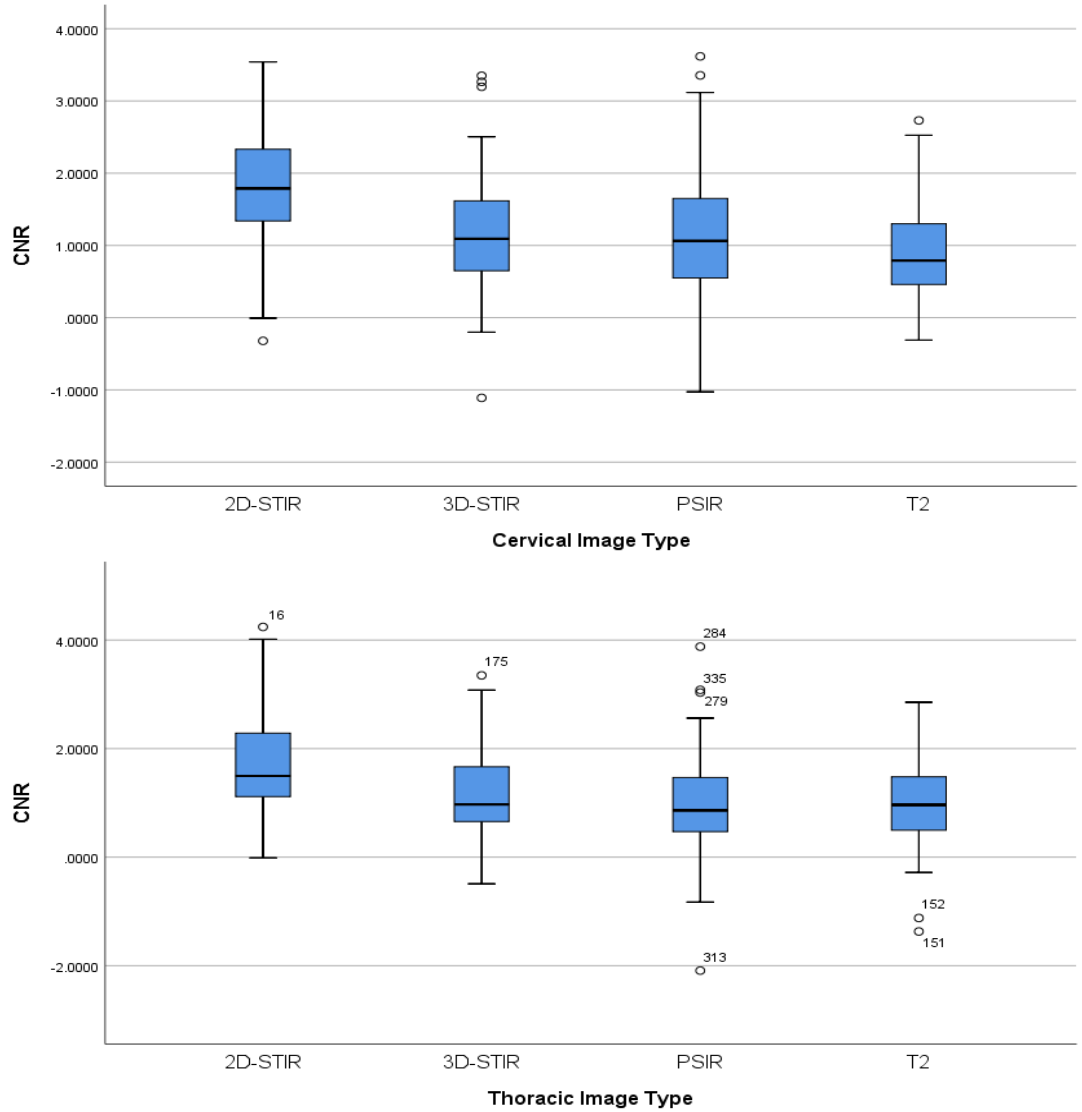

Figure 16: shows boxplots for CNR measurements. Black horizontal lines show the unadjusted average. There are some outliers, mostly in PSIR images and some CNR values below zero.

Additional Lesions

Data was sorted to investigate whether any other sequence from the initial evaluation detected lesions which were not detected on 2D-STIR. These are termed additional lesions. For each rater, a count of lesions not detected on 2D-STIR, but on another sequence was carried out. As this count includes false positives, the number of additional lesions which were confirmed by consensus reading was also counted in order to calculate PPV for additional lesions by rater. Findings are shown in Table 10.

Table 10: shows the number of additional lesions each sequence detected and how many of these were true lesions, according to consensus.

Cervical Images						
	Rater A			Rater B		
Weighting	**T2**	**3D-STIR**	**PSIR**	**T2**	**3D-STIR**	**PSIR**
No. Additional Lesions	11	32	15	10	10	7
No. True Additional Lesions	5	12	8	5	5	3
PPV of Additional Lesions	0.45	0.38	0.53	0.5	0.5	0.43
Thoracic Images						
	Rater A			Rater B		
Weighting	**T2**	**3D-STIR**	**PSIR**	**T2**	**3D-STIR**	**PSIR**
No. Additional Lesions	6	30	4	8	17	5
No. True Additional Lesions	6	15	4	4	3	3
PPV of Additional Lesions	1	0.5	1	0.5	0.18	0.6

For both raters and anatomical areas, the 3D-STIR weighting added the greatest number of lesions subsequently confirmed by consensus, but also a great number of false positives, as shown by the relatively low PPV. On thoracic images PPV was especially low for 3D-STIR by rater B (0.18). Rater A had a perfect PPV for additional lesions found on T2 and PSIR.

Data was subsequently sorted to count how many times a lesion was only seen on one sequence. This shows how often each weighting was solely responsible for lesion detection. The results are shown in Table 11.

Table 11 count of number of times a consensus confirmed lesion was only detected on one sequence by a given rater.

	Only seen on 2D-STIR and by Consensus		Only seen on T2 and by Consensus		Only seen on 3D-STIR and by consensus		Only seen on PSIR and by Consensus	
	Rater A	**Rater B**	**Rater A**	**Rater B**	**Rater A**	**Rater B**	**Rater A**	**Rater B**
Cervical	9	23	1	3	8	2	4	2
Thoracic	2	12	3	4	12	3	2	3

Data was analysed further in an attempt to characterise the size of true additional lesions not detected by either rater on 2D-STIR. The average size of the additional lesions found on cervical and thoracic scans is shown in Table 12. A simple independent samples t-test for significance was done to compare mean lesion size values of these additional lesions with all other lesions detected in the same anatomical scan area. Setting the level of significance to 0.05 shows additional lesions detected on the cervical scans were significantly smaller.

Table 12: average sizes of additional lesions vs those seen on 2D-STIR by at least one rater.

	Seen on 2D by at least one rater?	N	Mean Lesion Size (cm^2)	S.D	t-test p-value of difference
Cervical	**Yes**	98	0.287	0.102	0.021
	No	24	0.187	0.204	
Thoracic	**Yes**	67	0.31	0.21	0.21
	No	21	0.25	0.17	

Discussion

The weightings compared in this study were equalised in acquisition time, with all scans requiring roughly four minutes. This makes all weighting types comparable in terms of clinical usefulness per unit time. This is relevant in daily practice as scanner resources can be limited. An important consideration for radiological departments is: which scanning method gives the most diagnostic efficacy for a given amount of time? The results from this study should be able to help answer this question, as all sequences are approximately equal in scan time. A different approach would have been to equalise all sequences in resolution and SNR, in order to focus results solely on pathological contrast.

2D-STIR

2D-STIR was clearly the most diagnostically accurate sequence in the study. As well as achieving the best sensitivity for both raters, it also had a high PPV, indicating few false positives. Agreement was also substantial for 2D-STIR and compares favourably with all other sequences, being roughly equal to T2 and PSIR. This indicates that the sequence is reliable and consistent when used by different radiologists and for both cervical and thoracic images in this study. More lesions were found on 2D-STIR than any other sequence and this difference was statistically significant for both raters on cervical images and rater B on thoracic images.

STIR is no more prone to flow and motion artefacts than any other sequence including T2. Research has previously indicated that flow and motion artefacts seriously degraded sequence specificity. This meant it was not recommended to be used alone in MS lesion detection, but as an adjunct to at least one other sequence type less prone to artefacts (18). That 2D-STIR no longer seems affected by these artefacts most likely also contributes to the substantial level of agreement between radiologists and the +0.90 PPV. Technical explanations for this could possibly be due to the introduction of parallel imaging, which allows the number of excitations to be increased. This in turn suppresses the effect of motion artefacts and may also reduce flow artefacts as motion and flow may vary randomly between excitations (15). Another is the more time efficient implementation of FSE in inversion recovery. This is done by utilising two areas of “dead time” in the sequence. This is the time between inversion and excitation for a given slice, and also the end of readout and the start of the next repetition cycle. This dead time is used to invert, excite and readout other slices in the scan volume in 2D-imaging (21). The resultant shortening of acquisition time means more signal can be acquired by increasing number of excitations. Raters did not consider image noise as affecting detection: Noise was not a notable problem for 2D-STIR, to the likely benefit of sensitivity and PPV.

The success of 2D-STIR can primarily be attributed to the fact that 2D-STIR had a significantly better CNR than all other sequences. 2D-STIR exhibits a signal intensity difference between MS lesions and NASC which is markedly superior to all other sequences including PSIR, adjusted for noise. This is in line with previous studies, particularly for thoracic spinal cord imaging (34, 39, 41, 46). One notable

aspect is the generally much lower signal intensity of NASC on STIR. This became very apparent during the CNR measurement process in this study. NASC measurements were almost always lower for 2D-STIR in comparison with T2, whilst lesion intensity was often comparable. This suggests that the inversion pulse is having a good effect in terms of suppressing the fatty tissue of normally myelinated NASC, reducing its signal and improving CNR.

MS lesions lengthen both T1 and T2 times and increase proton density (7). It would seem likely that the 2D-STIR sequence takes advantage of more than just one of these effects, to the benefit of CNR. The sequence employed in this study had a fairly short TE of 38ms and a long TR of 3800ms. This gives an element of PD-weighting, whilst some T2 weighting is also present as the TE is not entirely short. The use of inversion recovery imbues an important element of T1 contrast, as tissues with shorter T1 times are darkened. The overall weighting is therefore a mixture of T1, T2 and PD. MS-lesions are known to vary in tissue content, having variations in the level of demyelination and axonal loss (6). Complete demyelination correlates with high intensity lesions and partial demyelination with intermediate intensity lesions. The mixture of weightings in 2D-STIR may be important for the detection of lesions where there is only partial demyelination, possibly explaining why 2D-STIR detected most lesions in this study.

But 2D-STIR did not detect all lesions. Sensitivity is far from 100% for any rater either cervically or thoracically. One likely reason is resolution. The results from this study show a difference in the sizes between lesions detected and those missed on 2D-STIR (shown on Table 12). For cervical images, lesion area was a third smaller for missed lesions and this difference was significantly different. The difference for thoracic images was not significant but showed the same tendency. That the difference was smaller on thoracic scans may hint that other factors may be involved. It is not likely to be CNR as these are roughly similar for 2D-STIR: 1.798 and 1.696 respectively.

However the high CNR and good ratings for image noise suggest that the sequence could be optimised further to detect smaller lesions by increasing resolution. Given that image resolution was lowest in the slice direction with a slice thickness of 2.5mm and a 10% slice gap, it would seem most appropriate to reduce slice thickness. The degree to which this can be done, whilst preserving the same scan time may however be limited. A marked reduction to 1mm for example, would very likely produce undiagnostic images full of noise. This would entail a signal reduction of 60%. Conceivably, a reduction to 2mm thickness would still produce favourable CNRs and PPV. But with this smaller change it is far from certain that more lesions would be detected.

3D-STIR

Image noise was an important factor limiting the efficacy of 3D-STIR. Both raters evaluated 3D-STIR as having significantly worse noise than all other sequences. An example of the images is shown in Figure 17. Images appeared grainy and of poor subjective quality to raters. This has led to some lesions being obscured, whilst simultaneously giving rise to false positives. The prevalence of false positives resulted in PPV being markedly worse on 3D-STIR compared to all other sequences, except for rater B on cervical images. Lower in-plane resolution in 3D-STIR compared to the other sequences may also have been a contributing factor. There was generally poor sensitivity for 3D-STIR compared to 2D-STIR and T2. 3D-STIR detected significantly fewer true MS lesions than 2D-STIR for both raters. 3D-STIR was in most situations more sensitive than PSIR, except for rater A on cervical scans. Difficulties with the efficacy of 3D-STIR is also reflected in the measures of agreement. 3D-STIR marks itself as being relatively unreliable and difficult to use by exhibiting a clearly lower kappa value of under 0.5. All the other sequences achieved a kappa of about 0.7 – 0.8.

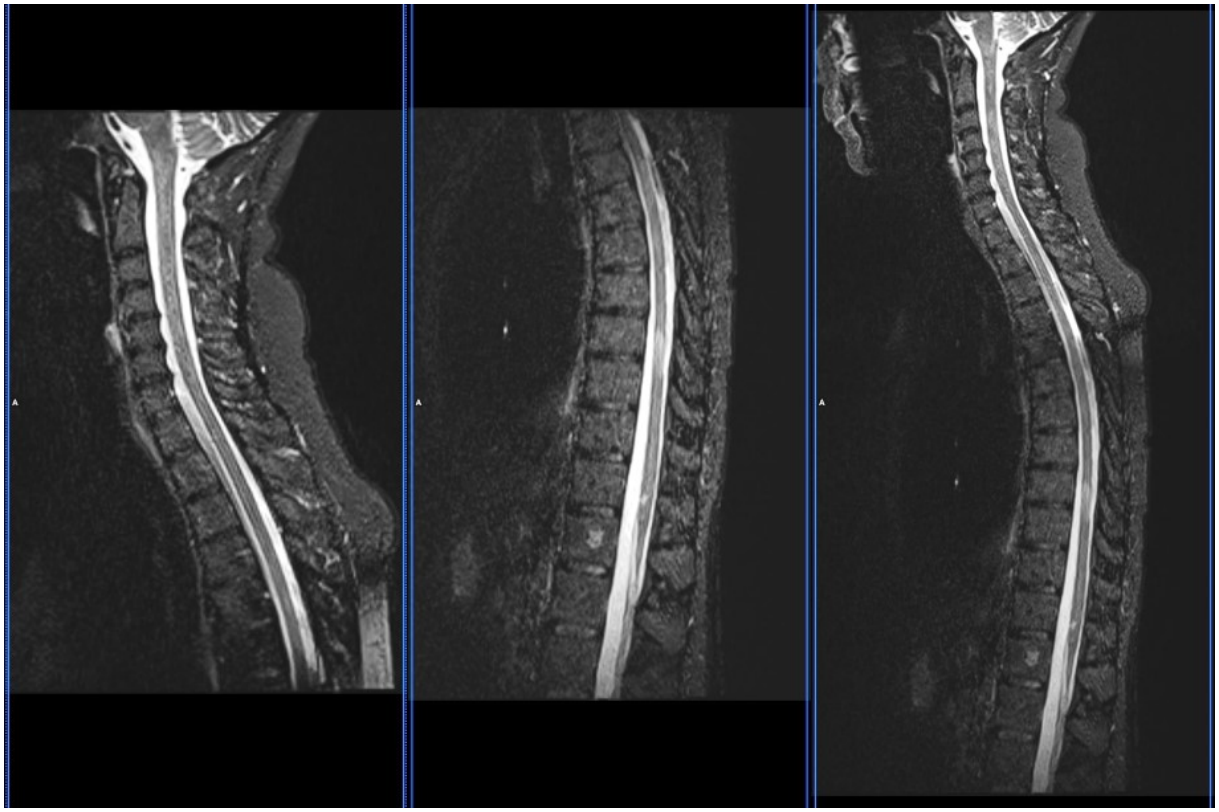

Figure 17: An example of the 3D-STIR sequence as acquired in this study. NASC has an inhomogeneous, grainy appearance.

Poorer lesion contrast compared to 2D-STIR could also explain some of the reduced efficacy. CNR is significantly worse than 2D-STIR. However, T2 had a better sensitivity for both raters and both cervically and thoracically, despite a poorer CNR. So even though lesion contrast is probably better on 3D-STIR than T2 it still detects fewer lesions, excluding CNR as an explanatory factor. It could be concluded that despite the use of variable angle refocusing pulses and appropriate k-space filling strategies, the 3D-FSE implementation of STIR in this study suffers from degraded lesion contrast relative to 2D acquisition. The chosen flip angle scheme was a T2 type. It was not possible to use a different one in this study's particular implementation of the sequence. Another flip angle scheme could perhaps be more MS lesion specific, for example a PD or T1 scheme, with a shorter effective TE, but this is not certain. The use of 3D gradient echo instead of FSE could also help address this problem, as contrast can be more readily controlled with more conventional parameters (flip-angle, TE, TR etc). The literature has four studies using sagittal 3D acquisition and three of these use gradient echo techniques. This was most often in conjunction with T1 weighting, often with PSIR reconstruction (33, 43, 49). This is a possible avenue of investigation at 1.5T.

Flow artefacts may also have affected diagnostic accuracy. These artefacts were judged significantly worse for 3D-STIR compared to all other sequences. Rater B considered this artefact as affecting diagnostic quality for sixteen patients. This artefact can obscure the delineation between spinal cord and CSF, as it causes a loss of signal from this fluid. It is noted that this sequence did not have flow compensation activated. This is because flow compensation can only reduce the effect of fluid flows in the readout or slice direction. For 3D-STIR, CSF flow along the spinal canal was in the in-plane phase direction, limiting the efficacy of flow compensation.

The thoracic 3D-STIR in particular seems to suffer from poor image quality, although subjective image quality was not evaluated separately from cervical scans in this study. The effect of the use of parallel imaging could be a contributing factor. For thoracic 3D scans, the arrangement of coils

possible in conjunction with the use of parallel imaging has detrimental effects. This is particularly due to the lack of an anterior coil. Less favourable geometric factors of the coil arrangement lead to more uncertainty in the image-unwrapping calculations of parallel imaging reconstruction (50). This can be seen as areas of increased noise in the image.

A few adjustments to sequence parameters may have helped increase SNR on 3D-STIR, whilst keeping scan times roughly the same. The TR of 3500 could have been reduced in a trade off with oversampling or increased number of excitations to increase signal. Another option could have been to reduce the acceleration factor for parallel imaging from three to two to reduce spatially variant noise. This increases with greater acceleration factors. This may have adversely affected noise in the spinal cord. Noise also increases when the geometry factor of the coil arrangement is of lower quality (29). This is especially relevant for the thoracic 3D-STIR where only posteriorly located coils were used. Whilst reducing acceleration factor does lengthen scan time, this could have been compensated for by reducing oversampling in the slice direction. This oversampling was perhaps unnecessary, as slab selective excitation was used. The number of slices could also have been reduced slightly. Increasing turbo factor could also have allowed the number of excitations to be increased without lengthening scan times. But this may have had a detrimental effect on lesion contrast, as effective TE is lengthened and blurring artefacts would possibly increase simultaneously.

When 3D acquisitions have been studied previously, it has always been done on 3T. Higher field strength has benefits for SNR, increasing signal in images. Additionally, acquisition times have generally been longer: in two studies, the 3D sequence involved had acquisitions times of 7.5 minutes. The performance of 3D-FSE sequence in this study seems to indicate that this sequence type suffers from ow image quality, when scan times of roughly four minutes are used at 1.5T. Higher field strength than 1.5T or longer scan times than four minutes appear necessary for reasonable image quality and resultant diagnostic efficacy.

In summary, 3D-STIR performed poorly in diagnostic accuracy. PPV, sensitivity and agreement are low due to poor subjective image quality.

Concomitant field effects?

The artefact, which gave an inhomogeneous appearance to some 3D-STIR acquisitions, is a curious effect. It may have reduced lesion detection in some cases, where the lesion was located near the crossover point between the cervical and thoracic images. Similar artefacts can often be seen on fat-suppressed images which use spectrally selective saturation techniques, due to patient induced variations in the static field. But this is unlikely for STIR which uses short inversion times instead to suppress fat. STIR is much less affected by inhomogeneity for this reason. Also, inhomogeneities are more usually seen at the nape of the neck, whereas the inhomogeneous areas on 3D-STIR were always in the corners of images.

One possibility is concomitant field effects, which can lead to phase errors and reduced image quality in FSE images (21). According to Maxwell's equations, when a gradient field is activated, it leads to higher order spatially variant perturbations in the overall magnetic field. These perturbations are non-linear. An example of the spatial effect of concomitant fields is shown in Figure 18. This is a sagittal phase image of a phantom from an article by Bernstein, Zhou (51). The pattern of the image disturbance bears a resemblance to the artefact found on 3D-STIR. Bernstein and Zhou note that these effects are usually unimportant, disturbing the evenness of the magnetic field by only >2ppm 20cm from the isocentre with gradient strengths of 10mT/m. The Siemens Avanto Fit system used for this study has a gradient strength of 45 mT/m and fields of view extend to around the same distance from the isocentre. 3D-FSE may require large gradient amplitudes in

order to reduce inter-echo times, meaning gradient amplitudes close to maximum may well have been employed in the sequence. This could perhaps have led to perturbations in the region of 10ppm. Bernstein and Zhou also note that concomitant field effects may be particularly problematic for 3D acquisitions, especially those with thick slabs of more than 20mm. For comparison, the slab thickness for 3D-STIR in this study was 64mm.

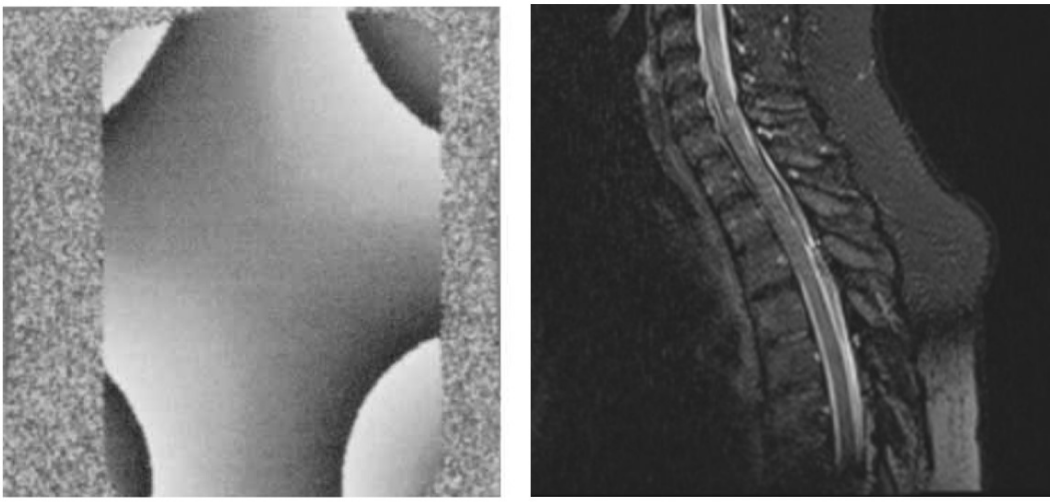

Figure 18: The left image shows the spatial variation in phase due to concomitant fields. Copied from Bernstein, Zhou (51). The pattern of effect on the image in the lower right corner bears a resemblance to the inhomogeneity artefact found on 3D-STIR, seen on the right.

Concomitant field effects can also lead to shifts in the apparent position of objects in the phase direction (52). It was noted during ROI measurement, that the copy-paste function for positioning ROI's often functioned sub-optimally with 3D-STIR. ROI's often had to be adjusted in position, even though they were pasted to the same organ position. This was the case for both in-plane and through-plane directions, both of which are spatially encoded with phase gradients. This is another indication that concomitant field effects are at play.

Intriguingly, this effect is also known to potentially reduce image quality by blurring details throughout the image and introducing particular ghosting effects. There are also indications that images can appear more noisy (52, 53). Image examples are included in the appendix. The artefact, if at work here, could be reducing lesion conspicuity by introducing blurring and increasing subjective noise. Reduction of the effect can be done by modifying hardware design, redesigning the pulse sequence, specifically gradient lobe shapes or by dedicated reconstruction methods (21). Possible parameter modifications could include using smaller fields-of-view or increasing bandwidth, although these would increase noise. The number of slices in the slab could also be reduced to decrease slab thickness. This also reduces signal but also acquisition time, so the number of excitations or over-sampling could be increased to compensate. This would seem a good option for reducing possible concomitant field effects.

T2

For both raters and anatomical areas T2 was less sensitive than 2D-STIR but more sensitive than 3D-STIR and PSIR. So T2 was second best in terms of sensitivity, despite a lower CNR than any other sequence. T2 found roughly just over half of all lesions detected by consensus. PPV was very good, at over 0.90 but not particularly better than 2D-STIR or PSIR. The primary reason for reduced sensitivity in comparison with 2D-STIR is likely the much reduced CNR T2 exhibits. This was significantly lower on T2 than for nearly all other sequences with the exception of the thoracic PSIR. This is most likely

because the sequence is primarily dependent on T2 changes in MS-lesions, and less so to T1 and perhaps PD changes. Whilst MS lesions do lengthen T2 times in spinal cord tissue, this effect is not great enough to produce CNRs comparable to 2D-STIR. Lesion dependent differences in proton density and T1 times are not fully utilised by this sequence and this hinders detection. At the same time and possibly as a consequence of this, signal from NASC on T2 is slightly higher, further reducing CNR.

For noise and artefacts, T2 performs well. This excludes these negative image effects as the reason for low sensitivity and is in line with T2's generally recognised ability to produce images relatively free from artefacts and of good subjective image quality.

In-plane resolution was relatively high for this sequence. Matrix sizes were 448 and 512 for the cervical and thoracic scans respectively giving a pixel size of approximately 0.6mm^2. It is therefore unlikely that lesions have been missed due to in-plane resolution. Perhaps if resolution in the slice direction had been improved, then this could have increased the number of detected lesions. This could easily be done by trading in-plane resolution for thinner slices. The number of phase encodings is reduced which reduces scan time whilst greater pixel area increases SNR, which could maintain signal and acquisition time despite the use of thinner slices. This may allow a reduction in slice thickness, possibly to below 2mm.

Sagittal T2 sequences are still recommended for use in spinal cord lesion detection as one of two included sequences (10, 19). This is because of the low prevalence of artefacts and good specificity. However in this study, 2D-STIR was equal to T2 in terms of artefacts, noise and PPV, so perhaps 2D-STIR can now be used alone, thereby shortening examination times.

One advantage of the inclusion of T2 is to allow better comparisons with other studies that include the same sequence. This furthers external generalizability.

PSIR

Sensitivity was low for PSIR. Rater B in particular found few lesions with this sequence with a sensitivity of about 37% in both anatomical areas. Rater A had more success with this sequence on cervical area images with a sensitivity of 53%. Agreement was on a par with 2D-STIR and T2 and PPVs were also very good, being over 0.90 for both raters and areas. PSIR actually achieved the highest PPV of all (0.97) when used by rater A in the thoracic area, with almost no false positives, so despite low sensitivity, PSIR does have good reliability.

The major apparent factor in low sensitivity is poorer CNR. This metric was comparable to 3D-STIR on cervical scans and T2 on thoracic. The poor CNR and low sensitivity found here is in contradiction with other studies investigating PSIR (33, 39, 46, 47). Two of these studies only compared PSIR with T2, but even so, T2 is clearly the more sensitive in the current paper. One important difference is perhaps that the other studies were carried out on 3T systems. That field strength is an important factor is supported by a recent study by Shayganfar et al at 1.5T, where PSIR was found to detect far fewer lesions than STIR (34). One possible criticism of that study is the rather long TE value of 70 ms on PSIR. This might have reduced the contribution of T1 signal in the image and therefore perhaps lesion conspicuity. More T1 signal would tend increase signal from NASC but leave signal from darker MS lesions unchanged or even decreased, improving contrast. However, in the current study, TE was much lower (15ms) and lesion detectability on PSIR was still poor.

Why PSIR lesion contrast at 1.5T should be lower than at 3T is unclear and rather intriguing. Whilst it is well known that T1 times vary with field strength, being shorter at lower values (54), the parameters of the PSIR sequence are adjusted to compensate. Inversion times in particular are

foreshortened both in this study and Shayganfar et al's work (TI=350ms), relative to 3T parameters (TI=400ms). Perhaps the spectral density function of MRI active protons in MS lesions is particularly suited in some way to imaging at 3T. This is very speculative. More likely, it could simply be that 3T's higher field strength is instrumental in increasing NASC signal to the extent that lesions appear relatively darker.

PSIR was less sensitive than T2 in this study, despite having a better CNR. The better CNR for PSIR could reasonably be expected to translate into better lesion contrast and conspicuity. This not found to be the case. More partial artefacts due to lower resolution could offer some explanation. The PSIR sequence in this study had the largest slice thickness of any sequence, possibly obscuring some lesions. A lateral lesion which borders onto CSF becomes more difficult to see due to this effect. The problem is exacerbated when slight spinal scoliosis causes the spinal cord to bend out of the imaging plane in the lateral direction. An example is shown in Figure 19. The thicker slices of PSIR could well have reduced the number of lesions detected.

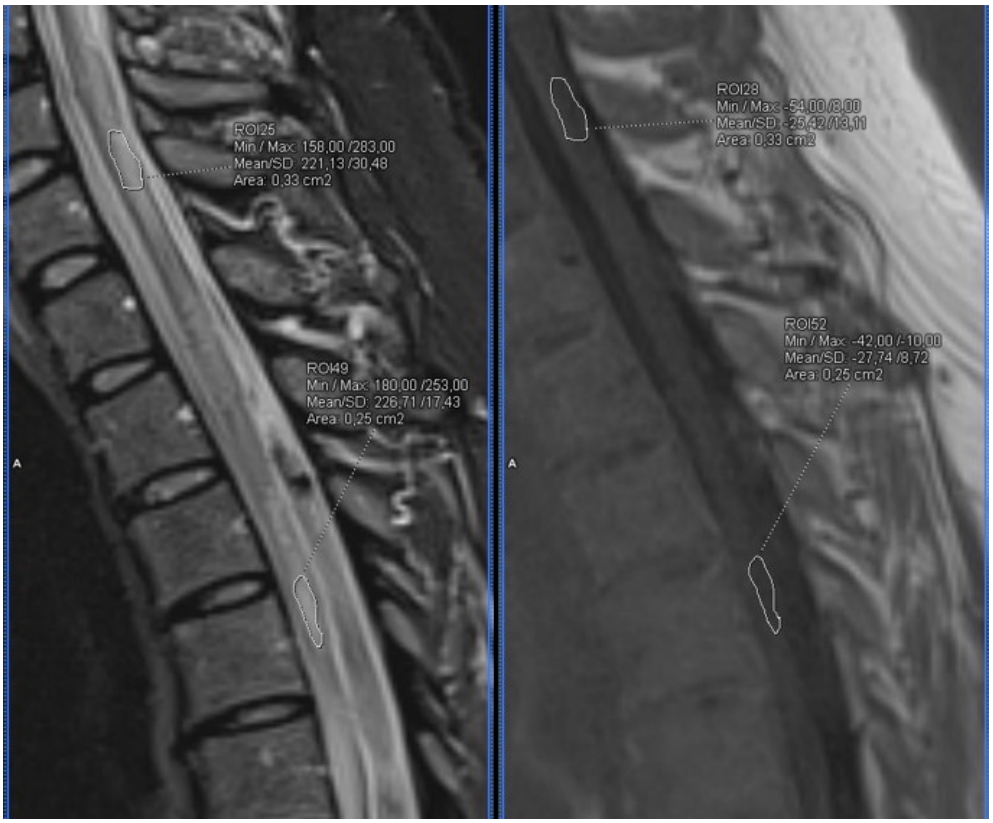

Figure 19: showing the effect of partial volume artefacts. The left image is 2D-STIR, the right image is PSIR. The lower lesion is more readily obscured due to the greater slice thickness of the PSIR sequence. Also, some CSF signal may have become included in the ROI.

Also, during ROI measurement it was noted that on a number of occasions, partial artefact effects could entail the inclusion of some CSF signal in the lesion ROI to an apparently greater extent than on other sequences with thinner slices. Whilst effort was made to avoid this where possible, in cases where it was difficult to delineate the lesion, some CSF may have been included anyway. As CSF intensities are very low, this would have the effect of erroneously increasing CNR for such lesions and skewing CNR results for PSIR. This is perhaps the case in Figure 19.

Previous studies used a form of CCR similar to equation 1 (46, 47), one of which also included CCR on magnitude reconstructed images of the PSIR sequence (47). Both studies found improved CCR for

PSIR and this concurred with better lesion detection. A third study used CNR in the form of equation 3, where PSIR was also found to be superior. Because a zero signal value on a PSIR image is not black – it is perfectly mid-grey - this can complicate comparisons with magnitude reconstructed sequences. CCR as calculated by equation 1 was considered and in fact calculated for use in this study, as it has been used in similar studies. However upon review of the results, the numbers seemed to indicate CCR for PSIR was extremely good – far better in fact than 2D-STIR by a factor of four. This simply did not concur with the appearance of lesions when compared between sequences. In this study, the average non-adjusted mean for NASC on both PSIR sequences was 5.3. For 2D-STIR the same value was 127. As described in the section on measuring contrast in diagnostic imaging studies, CCR measurements on PSIR, where NASC intensity values can be very small can be problematic, when being used to compare PSIR with magnitude reconstructed sequences. This could lead to speculation as to the interpretation of such comparisons in other studies which have used this metric (36, 46, 47). If conducted on scanners from other manufacturers, then perhaps signal intensities are not zero for mid-grey values on PSIR, as they are on Siemens scanners. They could conceivably be zero for absolute black instead, in the same manner as conventional magnitude images. This would make CCR comparisons valid. It is not known if this is the case.

Radiologists may perhaps find it psychovisually challenging to detect dark lesions on a mid-grey background, as opposed to bright lesions as with the other sequences: It may be more difficult to see dark lesions instead of bright lesions and raters were unaccustomed to lesion appearance on PSIR. This would naturally hinder lesion detection and skew results to the detriment of this sequence. Radiologist preference and the ingrained habit of looking for bright-on-dark could exaggerate this phenomenon. This is somewhat supported by the fact that radiologist A, (departmental radiologist) with fewer years experience was markedly better than radiologist B (consultant level) at detecting lesions on cervical scans. Rater A sensitivity was 53.7% (PPV 0.90) and rater B sensitivity was 37.7% (PPV 0.90). A simple way to counteract this effect would have been to include magnitude images in the evaluation as an adjunct to PSIR in the image evaluation. There is the risk of a reduction in contrast, when the magnetizations of lesions and NASC have different polarities, as shown in Figure 7, and this would have added an extra element to be accounted for. But as including these images requires no extra scan time, there would be few practical difficulties with their inclusion in clinical scanning. Shayganfar's work (34) indicates that magnitude images are better than PSIR reconstructed images, so their inclusion in the study would have been warranted. That study was published after all research scanning was completed and by that time magnitude reconstruction of the image data was no longer possible.

The use of window/level alterations during evaluation may also have had an effect. CNR measurements indicate a reasonable CNR on PSIR hinting that signal differences are there to be seen. It may however require judicious use of the windowing and level of image intensity display in order to make lesions more visible. Possible examples of this were noted during CNR measurement. Here, window/level was adjusted on some occasions in order to aid ROI placement. Figure 20 shows an example of a missed lesion, visible on PSIR but not on 2D-STIR. Altering windowing and level brings this likely lesion to the fore and may have been consistently utilised by raters. The use of windowing was not stipulated in the advice given to radiologists for image evaluation. As this is a normal part of radiological work, it was assumed that this would be done if the radiologist in question believed it necessary. It is therefore not known to what extent window/level was altered by raters during evaluation.

In the example mentioned above, the lesion in question was seen by both radiologists at the evaluation stage, but was notably missed at consensus. Windowing etc was not used extensively

during image review. This also brings into question the accuracy of the consensus process and highlights that it is an imperfect golden standard. Such missed lesions would of course lead to an underestimation of PSIR sensitivity and PPV and also leads to an over-estimation of the sensitivity of the other sequences including 2D-STIR.

Figure 20 shows a possible MS lesion missed by consensus. Far left is 2D-STIR, mid-left PSIR with unaltered window and level. Mid-right is a zoomed-in window and level adjusted PSIR image. Far right is the equivalent zoomed 2D-STIR. The red circle highlights an area, where a possible MS lesion can be seen as a darkened area in the windowed and zoomed PSIR image. No lesion can be seen on 2D-STIR. One vertebral level in the caudal direction, another lesion is clearly visible on both sequences and was detected by both raters and consensus.

As CSF is dark on PSIR, flow artefacts were minimal and indeed judged significantly better by rater A as they were rated as being absent. Some Flow artefacts were apparent according to rater B. They could be present due to some unwanted interaction between CSF flow and the inversion pulses. Partially ineffective inversion in combination with the excitation pulse could perhaps lead to a stimulated echo from CSF on readout. This could possibly happen if uninverted CSF from outside the slice flowed into the image before excitation. However, reduced levels of flow artefacts were not problematic for either 2D-STIR or T2, making the improvement on PSIR non-meaningful.

The potential for reducing slice thickness for PSIR in order to counteract partial artefacts seemed limited. In the pilot study, SNR was judged to generally be rather poor for PSIR, leading to the decision to use thicker slices. Perhaps it would have been a fairer comparison if this sequence had been adjusted to a thickness of 2.5mm to match 2D-STIR and T2. In order to increase signal to compensate for thinner slices, another parameter would require adjustment if SNR was to be maintained. It is difficult to see what this could be. In-plane resolution is one possibility but at the pilot study stage, the values used were considered the minimum: lower resolutions seemed of too poor quality. Another option could have been to increase turbo-factor in order to be able to trade off reduced scan time for increased signal, in order to reduce slice thickness. This was thought to have a negative effect on image contrast as it would increase the effective TE. This was considered undesirable as PSIR was thought of as a T1-weighted image, so this idea was not implemented. In hindsight this is considered a mistake, as other studies using PSIR in 2-dimensional form have employed turbo factors of seven.

Bias

Bias is a systematic influence on results which consistently leads to over- or underestimation of variables. It can limit the external validity and applicability of study results for example, the usefulness of this study in deciding which MRI sequence is most diagnostic (55). Reporting of possible sources of bias is necessary to ensure that results can be properly evaluated (56).

Image-based selection bias occurs when a particular modality or image type is used as an inclusion criteria. One inclusion criteria for patients in this study was that they were known to have spinal cord lesions. The reason was to maximise the overall number of lesions in the study and to improve experimental power in this respect. The department's MRI protocol used 2D-STIR and T2 images upon which lesions were originally detected, thus introducing the risk of image selection bias. This may have had the effect of excluding some patients, who may have lesions more readily detected on other sequences than STIR. A reasonable modification to the sampling method could have been to sample a random group of MS patients without reference to the presence of spinal cord lesions. Whilst it is always possible that the inclusion criteria may have skewed the lesions present in the cohort in 2D-STIRs favour, the risk of image-based selection bias is not considered not to be of overriding significance, as there are a large number of lesions in the sample group. It can be argued this should provide a broad spectrum of lesion types.

The study population should also reflect the proportion of disease sub-types if possible or focus on one sub type (56). The make-up of the study population roughly reflected the make-up of MS in general, with RRMS patients making up three quarters of the group. CIS and SPMS patients were also represented. All four subtypes were represented in the cohort. There was however only one PPMS patient. The relative lack number for this group could be a reflection of one inclusion criteria; it was a requisite that patients could stand. This could have excluded some potential wheel-chair participants, which may have excluded some PPMS sufferers. The reason for the inclusion criteria was that research scans were conducted out of hours with the author working alone. So there is perhaps an element of spectrum bias here, but the sample size is small and not grossly unrepresentative of MS patients as a whole. There are only three CIS patients and three SPMS in the cohort. This makes it difficult to draw conclusions and generalize about imaging findings for these small groups. Therefore, stratification of the data by MS type was avoided. Overall, the study population is roughly representative of MS patients as a whole. Results should therefore be generalizable in this manner.

Systematic bias in image acquisition was limited. The order in which different weightings were scanned was randomly chosen with a reasonable distribution of possible combinations. This was to mitigate any effects due to time dependent patient variables: lying still on an examination table can become more difficult as time progresses leading to motion artefacts.

Image Evaluation

Only two radiologists were involved in the study. This means less generalizability to the whole population of radiologists. Time and resource constraints were the main reason for this. However, there was a marked difference in experience between the two, with one radiologists being of consultant level in neuroradiology for over four years and the departmental radiologist having half a years experience as a specialist in neuroradiology.

Review bias can occur when raters are aware of results from other / competing diagnostic tests. Knowledge of patients clinical data or recall of details of their spinal cord scans from other image weightings will positively bias those scans which are evaluated last. In order to blind raters as effectively as possible, they were instructed there was to be a break of at least a few days between evaluations of different weightings. A week is recommended, but this was not specified. The order of patient presentation within each weighting was also randomised differently, with patients also having a different ID number in each image group.

Whilst rater preference may have biased results against PSIR, the opposite of this effect may also be at work to the advantage of T2. This sequence is very well understood and often used by the raters.

T2 is in fact the standard sequence at the department where the raters work, leading perhaps to a bias in favour of this sequence.

Reference Standard

The reference standard is of central importance to a diagnostic accuracy study (56). A fundamental difficulty is the lack of a true golden standard for the presence of a lesion. To do this truly accurately would require the pathological examination of all participants' spinal cords after scanning which is absolutely unacceptable. So consensus reviews must suffice even though it can be a poor substitute (56).

Consensus review found considerably more lesions than any single rater in the initial evaluation. Maximum sensitivity was achieved on 2D-STIR by rater B in the cervical area (see Table 6), but even here, some 32 lesions were not detected. So there is a notable difference between initial evaluations and consensus findings, requiring discussion of the reference standard.

As all four weightings were available to both raters at consensus, then the consensus opinion can be deemed as being reasonably acceptable and certainly superior to evaluations with a single sequence. Consensus review was also augmented by the use of axial reconstructions of the 3D-STIR sequence. This allowed spatial location of high intensity areas in the lateral direction, helping to determine whether partial artefacts were playing a part in (sagittal) image appearance, possibly adding to lesion detection where sagittal image findings were not deemed conclusive enough to indicate detection. This goes some way towards explaining why consensus review found roughly a third more lesions than any rater found using a single sequence. The inclusion of all four sequences most likely increased sensitivity by allowing the raters to double check all findings. Also on a practical level, the inclusion of more images of the same anatomy in the patient ensures that the given anatomy is evaluated more than once.

Another factor could be that limits for what could be deemed a lesion were somewhat different at consensus, especially with regard to size. It is possible that raters experienced a change in opinion as to how small a lesion can be or how it presents itself on a given image type. Generally, noise and graininess in images leads to scepticism from radiologists with regards to designating smaller details as pathology. This tendency has the effect of applying a threshold for how large a detail in the image must be, before it can be designated a lesion in the mind of the radiologist. This could have had an effect during initial evaluations, excluding smaller lesions. At consensus however, with access to more than one sequence, the threshold for the size of a detail in order to be designated lesion would be smaller. This is supported by the fact that the additional lesions found were smaller in size than those seen on 2D-STIR. Half of these additional lesions were only seen at consensus.

Variations on detection thresholds for consensus may also have occurred during the session also. Repetition of consensus review to evaluate concordance would have allowed the evaluation of possible variations.

Some of the above effect could have been investigated further by the use of visual grading analysis (VGA). Raters in this study used a dichotomous value for lesion detection – yes/no. Introducing a more gradual three or four point scale for likelihood of lesion presence would have been useful in this: definitely not – possible – likely – definite lesion. It would perhaps have led to raters to designate some smaller details as "possible" or "likely", which were subsequently confirmed at consensus. This would have revealed more information as to the nature of interpretation of some sequences. VGA studies also allow analysis of the receiver operating characteristic which may have given information on how large a detail must be in order to be reliably designated a lesion by raters.

There may also have been a certain element of context bias – or laboratory effect in consensus reading. This is the tendency to be too sensitive when taking part in a study. Under normal circumstances, radiologists can be more sceptical in relation to possible findings, as false positive diagnosis can very well lead to negative effects on patients. Under experimental conditions, there is no consequence for patients, which can lead to a marked oversensitivity (56). The consensus review may have seemed a more artificial situation than the initial evaluations, which perhaps more closely mimic normal clinical circumstances. It is possible that a certain eagerness at review may have led to some details being designated as lesions, when in fact they are not. Given the lack of a true golden standard this is very difficult to evaluate.

Rater preference for one of the sequences over another could also have biased the reference standard. Although the arbitrator attempted to ensure all images were reviewed, with many weightings available, each rater had a natural tendency to gravitate towards their "favourite" sequence, especially in the initial stages of evaluating a patient. These were the 2D- and 3D-STIR. Although it is natural to use 2D-STIR to obtain an overview of the patient given the superior CNR, there is the danger that some lesions have been overlooked because PSIR or T2 were not interrogated adequately during review. On at least one occasion, a lesion was seen at initial image evaluation of PSIR but not at consensus, as shown in Figure 20. If this error has occurred often, it would have the effect of overestimating 2D-STIR sensitivity, as many lesions not visible on this sequence, but perhaps visible on PSIR would be missed at consensus, due to the underutilisation of PSIR. But, there is reason to believe this bias is limited: PPV is high for both T2 and PSIR (PPV≥0.90). If many lesions were seen on PSIR but were erroneously missed at consensus, then PPV would be low for this sequence. The highest PPV was in fact achieved by PSIR (0.97) so lesions found during the initial evaluations were very much in agreement with consensus and few were therefore likely overlooked at the review stage. If many lesions had been seen on PSIR, but erroneously not at consensus, then PPV for PSIR would be lower.

Methodological Issues

The methodology of this study was not perfect. Intra-observer variation was not evaluated. The diagnostic accuracy of an observer is known to change over time. A measure of the variations in rater performance which can occur between viewings of the same images would have been useful to better evaluate the interactions between raters and the images (55). The stability of raters in detecting lesions could have had an important input to results, if for example, one type of sequence gave rise to a greater variation. It seems likely though, given the interrater agreements and noise measurements, that this would merely have revealed problems using the noisier 3D-STIR images and perhaps little else.

The artefact and noise evaluation was not divided by cervical and thoracic scans. It would have given a more precise evaluation of included sequences if this had been done. Noise was probably a greater problem for thoracic 3D-STIR. Separate subjective evaluations would have allowed this to be studied. Separate evaluations of motion artefacts could have revealed that motion due to swallowing was the only aspect affecting the artefact. If these had only been registered on cervical scans then this would have been a likely conclusion. Flow artefacts may also have affected cervical and thoracic images differently. It could also have been an advantage to specify pulsation artefacts as part of the subjective evaluation. These artefacts are known to be able to cause problems in spinal cord imaging. Pulsation was not noted during scanning or review however. Also, it is reasonable to assume that the effect of this artefact would have manifested itself in the evaluations of motion, which was rarely problematic for raters.

Spine labelling was not used. This could have helped avoid some of the issues with lesion location which occurred in this study. It was attempted to do so after image acquisition, but there were significant problems with exporting labelling data with images onto disc media, as part of the preparations for evaluation.

The potential of 3D-STIR and PSIR was not fully utilised in this study. Magnitude images were not produced and used for PSIR. Axial reconstruction of the 3D-STIR was also easy to implement for consensus and could have been used at the initial evaluation stage. In retrospect, it is considered it would have been appropriate to produce these images as an adjunct to the native sagittal images for evaluation.

The relative SNRs of sequences were not measured in this study. This could have been done with a phantom and ROI measurements prior to scanning of patients and would have enabled comparison of overall SNR between sequences and adjustment to make them more equal. Although the process is involved, especially when parallel imaging is used, it could potentially have helped reveal certain large differences in SNR prior to imaging of patients. More equal SNR between weightings would have focused results more on the effect of resolution and contrast.

The sequences studied in this paper are not the only possibilities for spinal cord imaging in MS. Most notably, the recommended dual-echo sequence was not included. Proton density weighted images and full length axial T2 images have also been studied but were not included. There are also the more experimental types of images. Quantitative MRI, magnetization transfer imaging, myelin water fraction imaging, diffusion and diffusion tensor imaging have potential as research tools. Pathological mechanisms and tissue cellularity can be examined with these sequences. One or more may come into regular clinical use in relation to MS at some point.

Detecting Lesions

It is important to diagnose MS as early as possible to enable modification of disease course and to aid in prognosis. Sensitive detection of MS lesions in the spinal cord is an important contributory factor in this (12, 13). Various tissue changes as a result of MS lead to signal changes in MRI parameters. These are demyelination, axonal loss, inflammation, gliosis and blood-brain-barrier damage resulting in a heterogeneous appearance of lesions on MRI. It is difficult to predict which type of tissue changes an undiagnosed MS patient may present. A lesion can be made up of one or a combination of these and multiple lesions within a patient will vary.

Sensitive detection of lesions is dependent on a good CNR response to each type of tissue change and different sequence types respond differently. Whilst results from post-mortem studies of tissue changes should be interpreted with caution, correlations between various MRI parameters, such as proton density, T1- and T2-times have been investigated. This can perhaps shed some light on why some sequences have more CNR than others. One post-mortem study showed that partial demyelination and a penumbra of inflammation can be visualised on MRI when PD-weighting is employed. Inflammatory tissue changes are seen as medium intensity lesions. At ultra-high field with excised spinal cords, PD images revealed inflammation more clearly than pathological examination, suggestive that this weighting is useful for visualising this type of tissue change (6). Another post-mortem study on non-fixed spinal cord samples from three MS patients found that demyelination correlated most strongly with T1 lengthening and axonal loss correlated most strongly with increased proton density. These tissue changes also correlated with lengthened T2 times, but to a lesser degree (8). So different types of tissue changes primarily change different MRI parameters. One correlates best with T1 lengthening, the other an increase in proton density and T2 changes correlate less strongly.

2D-STIR has cemented its position as the most sensitive to MS lesions in this study. This is in line with a number of peer-reviewed articles which have found good results for 2D-STIR (34, 39, 41, 46, 47). It would appear that 2D-STIR responds more favourably to the variety of these changes in comparison with other sequences. 2D-STIR as implemented here has elements of T1, T2 and PD weighting, all of which are to varying extents correlates of the various MS induced tissue changes. This is perhaps allowing the sequence to visualise a greater variety of MS lesions. PSIR CNR is likely dependent on changes in T1 time more than any other MRI parameter and the equivalent can be said of T2 images. This may be limiting CNR for one or more types of MS tissue changes for these sequences, reducing detection sensitivity. It seems likely that 2D-STIR has a more favourable signal response to more changes in MRI parameters due to MS lesions. This is the most likely explanation for higher sensitivity of in this and other studies.

Resolution is likely the limiting factor for 2D-STIR in detecting more lesions. In the planning stages of this study, it appeared 2D-STIR had an advantage in terms of CNR at 1.5T, so it seemed logical to pursue a 3D-STIR solution. However, the implementation here seems impractical due to poor subjective image quality, leading to low PPV and agreement. CNR was unremarkable. Full length axial scanning is yet to be studied at 1.5T. It has the effect of increasing resolution and improves lesion detection of smaller lesions and laterally placed lesions (35, 37). This is achievable in under three minutes and superior to sagittal T2 at 3T. It would have been interesting to see if the same could be done at 1.5T in comparison with 2D-STIR, possibly with optimised PD-weighted axial images. At some point in the future, simultaneous multi-slice MRI with FSE may come into regular clinical use. This technique is capable of reducing scan times by acquiring image data from different slices at the same time. This technique has been applied to 2D-FSE sequences (57) and could possibly allow full axial coverage of the spinal cord with short acquisition times (<2 minutes) at 1.5T with good SNR. This could improve the viability of full-coverage axial scanning of the spinal cord in MS.

Best Combination of sequences

The results of this study indicate 2D-STIR is recommendable as the primary sequence for MS lesion detection in the spinal cord. Both MAGNIMS and the International Panel for MS recommend the use of two sequences as it is not certain that any single sequence can be used alone. The fact that consensus, which used four sequences to detect lesions, found a number of lesions not seen on any single sequence supports the use of multiple sagittal sequences. Study findings are analysed here to investigate which sequence can be recommended in conjunction with 2D-STIR, which had the highest sensitivity.

It would be reasonable to theorise that one type of lesion, perhaps a lesion consisting mostly of axonal loss, would be particularly conspicuous on a particular sequence – for example PSIR. But the overriding majority of lesions were seen on more than one sequence in this study. Only few lesions were detected on a single sequence. This is shown in Table 11 for each rater. The number of lesions for which this was the case gives an indication of a sequence's unique ability to detect lesions in a manner different from other sequences. That 2D-STIR was more sensitive is reflected in the generally greater number of lesions detected alone on this sequence. Otherwise, the numbers are relatively low for the other sequences. This indicates that although lesions are heterogeneous, no single sequence type other than 2D-STIR had a unique ability to detect a specific subset of lesions, as confirmed by consensus. Therefore, all included sequences generally detect lesions in a similar manner, but merely with different sensitivities. This is on the proviso that consensus did not systematically miss lesions for a given sequence.

Analysis of the data in Table 10 shows how many additional lesions were found by each rater on sequences other than 2D-STIR. This indicates how much added sensitivity a given sequence may potentially offer. The generally low PPV for additional lesions is problematic and points towards non-concordance problems. It is difficult to conclude whether this is due to misinterpretation at the initial evaluation, or whether there was a marked difference in interpretation thresholds at consensus. But analysis of additional lesions is a useful guide to choosing the best sequence to accompany 2D-STIR.

For rater A in the cervical area: if the number of true lesions found on 3D-STIR are added to those found on 2D-STIR, then sensitivity increases from 68%0 to 78.8% with the additional detection of twelve further lesions. The PPV of additional lesions seen on 3D-STIR was low at 0.38: 3D-STIR gives rise to many false positives, but this could be offset by its use in conjunction with 2D-STIR. PSIR also adds a notable number of lesions for rater A and would increase sensitivity to about 75%. Whilst overall PPV on PSIR for was high for rater A (PPV=0.90) PPV for additional lesions is low (PPV=0.53), but perhaps this is an error due to the underutilisation of PSIR at consensus. So on cervical scans, there is some ambiguity as to whether PSIR or 3D-STIR adds the most benefit for rater A. In the thoracic area, 3D-STIR adds fifteen lesions and sensitivity thereby increases from 53.4% to 67%, again increasing rater A's sensitivity towards that of rater B. But worryingly, seven of these additional lesions are from a single SPMS patient, giving some suspicion of reader variation during consensus. Subtracting these seven lesions makes 3D-STIR rather equal to T2 in number of additional lesions in the thoracic area for rater A. The PPV of additional lesions was perfect at 1.0 for both T2 and PSIR, whereas 3D-STIR was 0.5. Given that T2 found more additional lesions and that previous studies have found PSIR to underperform on thoracic scans, then T2 seems to give the most overall benefit to rater A.

For rater B in the cervical area both T2 and 3D-STIR add five lesions increasing sensitivity from 73.8% to 77.9%. PPV for both sequence's additional lesions was 0.5. PSIR only added three lesions. This makes either T2 or 3D-STIR recommendable on cervical scans. On thoracic area scans, rater B detected only few additional lesions, with T2 adding the most with four lesions. This would increase sensitivity from 71.6% to 76.1%. Interestingly, 82% of rater B's additional lesions on 3D-STIR were not confirmed at consensus (PPV=0.18). It seems unlikely this is purely due to poor image quality leading to false positives. It seems more likely that an element of intra- and/or inter-reader variation is again playing a role. The use of a VGA method to investigate thresholds for the designation of a lesion and measures of intrareader concordance could have been useful in the analysis of this result. But it is clear that poor image quality for 3D-STIR has had a great effect, leading to low PPV. Given that T2 added the greatest (though only four) lesions and that this sequence is robust, T2 adds the greatest overall benefit for rater B on thoracic scans.

Although cervical and thoracic scans were not evaluated separately for image quality by raters, it is very likely that the thoracic 3D-STIR was particularly poor, as PPVs were especially low. 3D-STIR in this area is therefore unreliable. That 3D-STIR works better on cervical scans than thoracic is not surprising. Improved geometrical factors improve the function of parallel imaging and possible concomitant field effects are lesser, as the field-of-view is smaller.

Analysis of additional lesions shows sensitivity for rater A increases most when 2D-STIR is accompanied by either 3D-STIR or PSIR. For rater B, either T2 or 3D-STIR would be a reasonable choice. Despite the fact that 3D-STIR has low sensitivity, PPV and poor image quality, it seems to contribute most to the number of lesions detected, when used together with 2D-STIR. Perhaps this is due to thinner slices. T2 seems to be the most sensible choice for thoracic scans for both raters, when taking possible experimental error and 3D-STIR image quality in this area into account. PSIR

appears very insensitive thoracically due to poor contrast and this is in agreement with a study by Alcaide-Leon et al (39).

Using 3D-STIR in conjunction with 2D-STIR for scanning the cervical spinal cord would however result in rather lengthy examination times as this requires at least four minutes. There are also indications that 3D-STIR would be of lesser help to experienced radiologists. For most departments, it may seem a too time-costly sequence with rather poor image quality to be a welcome addition. Good quality T2 images on the other hand, can be acquired relatively quickly. Using the sequence in this study as a starting point, in-plane resolution can be reduced, possibly allowing both slice-thickness and acquisition time to be reduced, whilst preserving a reasonable contribution to diagnostic efficacy. The sequence is robust, fairly sensitive, well understood and widely used in radiology. This would also tally well with the choice of T2 sequences as the second sequence for thoracic scans.

Conclusion

Early-as-possible diagnosis in the initial stages of MS and accurate monitoring of progression are important aspects of the diagnostic efficacy of spinal cord MRI imaging. The central aspect of this is lesion detection. In answer to the research question: 2D-STIR was the most diagnostically accurate sequence included in this study. Both raters detected significantly more MS lesions than all other sequences for cervical scans with this sequence. One rater detected significantly more lesions on 2D-STIR thoracic scans. Contrast ratios for 2D-STIR were significantly greater than that of all other sequences for both cervical and thoracic scans. Sensitivity was superior for 2D-STIR consequently. There were no great differences in PPV between three of the four sequences included in this study. Only 3D-STIR suffered from a markedly lower PPV. Interrater agreement was similar for all sequences except 3D-STIR, which again was markedly worse. Agreement and PPV on 2D-STIR were found comparable to T2 indicating that the reliability of 2D-STIR has improved. Problems with image quality, which have previously meant STIR was not recommended to be used alone (18), are now much less apparent.

Guidelines still recommend the use of at least two weightings for spinal cord MRI of MS patients. Which sequence contributes most benefit as the second weighting with 2D-STIR depends on scan area – cervical or thoracic – and may also be dependent on experience. 3D-STIR has potential for cervical scans, perhaps mostly for less experienced radiologists. But this is unlikely the case for thoracic scans. T2 was still found to be the second most sensitive sequence in the study, was reliable and generally contributed consistently to lesions not seen on 2D-STIR. T2 was not degraded by its use in thoracic scanning, making it a more logical choice in conjunction with imaging of the whole spinal cord.

There are some issues with method in this study. Problems with the imperfect reference standard, possible rater preference for images, image-based selection bias and possible reader variations reduce validity somewhat. Despite this, this study has confirmed the recommendability of 2D-STIR for MS lesions detection, especially at 1.5T field strength. Given equal acquisition times, 2D-STIR is a sensitive and reliable sequence, with a low prevalence of flow and motion artefacts. 2D-STIR did not find all lesions, seemingly missing smaller lesions, suggesting resolution as a limiting factor. PSIR underperformed due to poor lesion contrast, solidifying evidence that this sequence is for some reason less useful at 1.5T. 3D-STIR underperformed primarily due to image noise and generally poorer image quality but may be more sensitive to small lesions. T2 is a reliable choice as the supplementary weighting, although there are indications that 2D-STIR may now be reliable enough to be used alone.

References

1. MASSIMO FILIPPI PP, AND MARIA A. ROCCA. Neuroimaging. Part 1: Elsevier; 2016.
2. Daroff RB, Bradley WG. Bradley's neurology in clinical practice / [edited by] Robert B. Daroff ... [et al.]. 6th ed. Philadelphia, PA: Elsevier/Saunders; 2012.
3. Filippi M, Bruck W, Chard D, Fazekas F, Geurts JJG, Enzinger C, et al. Association between pathological and MRI findings in multiple sclerosis. Lancet Neurol. 2019;18(2):198-210.
4. Schmierer K, McDowell A, Petrova N, Carassiti D, Thomas DL, Miquel ME. Quantifying multiple sclerosis pathology in post mortem spinal cord using MRI. Neuroimage. 2018;182:251-8.
5. Brownlee WJ, Hardy TA, Fazekas F, Miller DH. Diagnosis of multiple sclerosis: progress and challenges. Lancet. 2017;389(10076):1336-46.
6. Bot JC, Blezer EL, Kamphorst W, Lycklama ANGJ, Ader HJ, Castelijns JA, et al. The spinal cord in multiple sclerosis: relationship of high-spatial-resolution quantitative MR imaging findings to histopathologic results. Radiology. 2004;233(2):531-40.
7. Nijeholt GJ, Bergers E, Kamphorst W, Bot J, Nicolay K, Castelijns JA, et al. Post-mortem high-resolution MRI of the spinal cord in multiple sclerosis: a correlative study with conventional MRI, histopathology and clinical phenotype. Brain. 2001;124(Pt 1):154-66.
8. Mottershead JP, Schmierer K, Clemence M, Thornton JS, Scaravilli F, Barker GJ, et al. High field MRI correlates of myelin content and axonal density in multiple sclerosis--a post-mortem study of the spinal cord. J Neurol. 2003;250(11):1293-301.
9. McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol. 2001;50(1):121-7.
10. Filippi M, Rocca MA, Ciccarelli O, De Stefano N, Evangelou N, Kappos L, et al. MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. Lancet Neurol. 2016;15(3):292-303.
11. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 2018;17(2):162-73.
12. Kearney H, Miller DH, Ciccarelli O. Spinal cord MRI in multiple sclerosis--diagnostic, prognostic and clinical value. Nat Rev Neurol. 2015;11(6):327-38.
13. Arrambide G, Rovira A, Sastre-Garriga J, Tur C, Castillo J, Rio J, et al. Spinal cord lesions: A modest contributor to diagnosis in clinically isolated syndromes but a relevant prognostic factor. Mult Scler. 2018;24(3):301-12.
14. Zackowski KM, Smith SA, Reich DS, Gordon-Lipkin E, Chodkowski BA, Sambandan DR, et al. Sensorimotor dysfunction in multiple sclerosis and column-specific magnetization transfer-imaging abnormalities in the spinal cord. Brain. 2009;132(Pt 5):1200-9.
15. Westbrook C, Roth CK, Talbot J. MRI in practice. 3rd ed. Malden, MA: Blackwell Pub.; 2005. xii, 410 p. p.
16. McRobbie DW. MRI from picture to proton. 2nd ed. Cambridge, UK ; New York: Cambridge University Press; 2007. xii, 394 p. p.
17. Gilmore CP, Geurts JJ, Evangelou N, Bot JC, van Schijndel RA, Pouwels PJ, et al. Spinal cord grey matter lesions in multiple sclerosis detected by post-mortem high field MR imaging. Mult Scler. 2009;15(2):180-8.
18. Bot JC, Barkhof F, Lycklama a Nijeholt GJ, Bergers E, Polman CH, Ader HJ, et al. Comparison of a conventional cardiac-triggered dual spin-echo and a fast STIR sequence in detection of spinal cord lesions in multiple sclerosis. Eur Radiol. 2000;10(5):753-8.
19. Traboulsee A, Simon JH, Stone L, Fisher E, Jones DE, Malhotra A, et al. Revised Recommendations of the Consortium of MS Centers Task Force for a Standardized MRI Protocol and Clinical Guidelines for the Diagnosis and Follow-Up of Multiple Sclerosis. AJNR Am J Neuroradiol. 2016;37(3):394-401.

20. Rovira A, Wattjes MP, Tintore M, Tur C, Yousry TA, Sormani MP, et al. Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis-clinical implementation in the diagnostic process. Nat Rev Neurol. 2015;11(8):471-82.
21. Bernstein MA, King KF, Zhou ZJ. Handbook of MRI pulse sequences. Amsterdam ; Boston: Academic Press; 2004. xxii,1017 p. p.
22. Hou P, Hasan KM, Sitton CW, Wolinsky JS, Narayana PA. Phase-sensitive T1 inversion recovery imaging: a time-efficient interleaved technique for improved tissue contrast in neuroimaging. AJNR Am J Neuroradiol. 2005;26(6):1432-8.
23. Smith SA, Edden RA, Farrell JA, Barker PB, Van Zijl PC. Measurement of T1 and T2 in the cervical spinal cord at 3 tesla. Magn Reson Med. 2008;60(1):213-9.
24. Borrello JA, Chenevert TL, Aisen AM. Regional phase correction of inversion-recovery MR images. Magn Reson Med. 1990;14(1):56-67.
25. Xiang QS. Inversion recovery image reconstruction with multiseed region-growing spin reversal. J Magn Reson Imaging. 1996;6(5):775-82.
26. Mugler JP, 3rd. Optimized three-dimensional fast-spin-echo MRI. J Magn Reson Imaging. 2014;39(4):745-67.
27. Grussu F, Schneider T, Zhang H, Alexander DC, Wheeler-Kingshott CA. Neurite orientation dispersion and density imaging of the healthy cervical spinal cord in vivo. Neuroimage. 2015;111:590-601.
28. Yiannakas MC, Grussu F, Louka P, Prados F, Samson RS, Battiston M, et al. Reduced Field-of-View Diffusion-Weighted Imaging of the Lumbosacral Enlargement: A Pilot In Vivo Study of the Healthy Spinal Cord at 3T. PLoS One. 2016;11(10):e0164890.
29. Dietrich O, Raya JG, Reeder SB, Reiser MF, Schoenberg SO. Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. J Magn Reson Imaging. 2007;26(2):375-85.
30. Gringel T, Schulz-Schaeffer W, Elolf E, Frolich A, Dechent P, Helms G. Optimized high-resolution mapping of magnetization transfer (MT) at 3 Tesla for direct visualization of substructures of the human thalamus in clinically feasible measurement time. J Magn Reson Imaging. 2009;29(6):1285-92.
31. Prados F, Ashburner J, Blaiotta C, Brosch T, Carballido-Gamio J, Cardoso MJ, et al. Spinal cord grey matter segmentation challenge. Neuroimage. 2017;152:312-29.
32. Karavasilis E, Velonakis G, Argiropoulos G, Athanasakos A, Poulou LS, Toulas P, et al. Proton Density Fat Suppressed MRI in 3T Increases the Sensitivity of Multiple Sclerosis Lesion Detection in the Cervical Spinal Cord. Clin Neuroradiol. 2019;29(1):45-50.
33. Fechner A, Savatovsky J, El Methni J, Sadik JC, Gout O, Deschamps R, et al. A 3T Phase-Sensitive Inversion Recovery MRI Sequence Improves Detection of Cervical Spinal Cord Lesions and Shows Active Lesions in Patients with Multiple Sclerosis. AJNR Am J Neuroradiol. 2019;40(2):370-5.
34. Shayganfar A, Sarrami AH, Fathi S, Shaygannejad V, Shamsian S. Phase sensitive reconstruction of T1-weighted inversion recovery in the evaluation of the cervical cord lesions in Multiple Sclerosis; is it similarly eligible in 1.5 T magnet fields? Mult Scler Relat Disord. 2018;23:17-22.
35. Breckwoldt MO, Gradl J, Hahnel S, Hielscher T, Wildemann B, Diem R, et al. Increasing the sensitivity of MRI for the detection of multiple sclerosis lesions by long axial coverage of the spinal cord: a prospective study in 119 patients. J Neurol. 2017;264(2):341-9.
36. Sundarakumar DK, Smith CM, Hwang WD, Mossa-Basha M, Maravilla KR. Evaluation of Focal Cervical Spinal Cord Lesions in Multiple Sclerosis: Comparison of White Matter-Suppressed T1 Inversion Recovery Sequence versus Conventional STIR and Proton Density-Weighted Turbo Spin-Echo Sequences. AJNR Am J Neuroradiol. 2016;37(8):1561-6.
37. Galler S, Stellmann JP, Young KL, Kutzner D, Heesen C, Fiehler J, et al. Improved Lesion Detection by Using Axial T2-Weighted MRI with Full Spinal Cord Coverage in Multiple Sclerosis. AJNR Am J Neuroradiol. 2016;37(5):963-9.

38. Chong AL, Chandra RV, Chuah KC, Roberts EL, Stuckey SL. Proton Density MRI Increases Detection of Cervical Spinal Cord Multiple Sclerosis Lesions Compared with T2-Weighted Fast Spin-Echo. AJNR Am J Neuroradiol. 2016;37(1):180-4.
39. Alcaide-Leon P, Pauranik A, Alshafai L, Rawal S, Oh J, Montanera W, et al. Comparison of Sagittal FSE T2, STIR, and T1-Weighted Phase-Sensitive Inversion Recovery in the Detection of Spinal Cord Lesions in MS at 3T. AJNR Am J Neuroradiol. 2016;37(5):970-5.
40. Riederer I, Karampinos DC, Settles M, Preibisch C, Bauer JS, Kleine JF, et al. Double inversion recovery sequence of the cervical spinal cord in multiple sclerosis and related inflammatory diseases. AJNR Am J Neuroradiol. 2015;36(1):219-25.
41. Nayak NB, Salah R, Huang JC, Hathout GM. A comparison of sagittal short T1 inversion recovery and T2-weighted FSE sequences for detection of multiple sclerosis spinal cord lesions. Acta Neurol Scand. 2014;129(3):198-203.
42. Ozturk A, Aygun N, Smith SA, Caffo B, Calabresi PA, Reich DS. Axial 3D gradient-echo imaging for improved multiple sclerosis lesion detection in the cervical spinal cord at 3T. Neuroradiology. 2013;55(4):431-9.
43. Nair G, Absinta M, Reich DS. Optimized T1-MPRAGE sequence for better visualization of spinal cord multiple sclerosis lesions at 3T. AJNR Am J Neuroradiol. 2013;34(11):2215-22.
44. Martin N, Malfair D, Zhao Y, Li D, Traboulsee A, Lang D, et al. Comparison of MERGE and axial T2-weighted fast spin-echo sequences for detection of multiple sclerosis lesions in the cervical spinal cord. AJR Am J Roentgenol. 2012;199(1):157-62.
45. White ML, Zhang Y, Healey K. Cervical spinal cord multiple sclerosis: evaluation with 2D multi-echo recombined gradient echo MR imaging. J Spinal Cord Med. 2011;34(1):93-8.
46. Philpott C, Brotchie P. Comparison of MRI sequences for evaluation of multiple sclerosis of the cervical spinal cord at 3 T. Eur J Radiol. 2011;80(3):780-5.
47. Poonawalla AH, Hou P, Nelson FA, Wolinsky JS, Narayana PA. Cervical Spinal Cord Lesions in Multiple Sclerosis: T1-weighted Inversion-Recovery MR Imaging with Phase-Sensitive Reconstruction. Radiology. 2008;246(1):258-64.
48. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
49. Mirafzal S, Goujon A, Deschamps R, Zuber K, Sadik JC, Gout O, et al. 3D PSIR MRI at 3 Tesla improves detection of spinal cord lesions in multiple sclerosis. J Neurol. 2019.
50. Larkman DJ, Nunes RG. Parallel magnetic resonance imaging. Phys Med Biol. 2007;52(7):R15-55.
51. Bernstein MA, Zhou XJ, Polzin JA, King KF, Ganin A, Pelc NJ, et al. Concomitant gradient terms in phase contrast MR: analysis and correction. Magn Reson Med. 1998;39(2):300-8.
52. Weavers PT, Tao S, Trzasko JD, Frigo LM, Shu Y, Frick MA, et al. B0 concomitant field compensation for MRI systems employing asymmetric transverse gradient coils. Magn Reson Med. 2018;79(3):1538-44.
53. Zhou XJ, Tan SG, Bernstein MA. Artifacts induced by concomitant magnetic field in fast spin-echo imaging. Magn Reson Med. 1998;40(4):582-91.
54. Bottomley PA, Foster TH, Argersinger RE, Pfeifer LM. A review of normal tissue hydrogen NMR relaxation times and relaxation mechanisms from 1-100 MHz: dependence on tissue type, NMR frequency, temperature, species, excision, and age. Med Phys. 1984;11(4):425-48.
55. Portney LG, Watkins MP. Foundations of clinical research : applications to practice. 3rd edition, revised . ed. Upper Saddle River, N.J.: Pearson/Prentice Hall; 2015. xix, 892 pages p.
56. Gennaro G. The "perfect" reader study. Eur J Radiol. 2018;103:139-46.
57. Gagoski BA, Bilgic B, Eichner C, Bhat H, Grant PE, Wald LL, et al. RARE/turbo spin echo imaging with Simultaneous Multislice Wave-CAIPI. Magn Reson Med. 2015;73(3):929-38.

Appendices

Appendix 1: Full list of parameters for all image sequences.

SIEMENS MAGNETOM Avanto_fit

Table of contents

\\USER

PROJEKTER

pwal

Medulla Projekt

I_AASpine_Scout	*
II_AASpine_Scout	*
I_t2_tse_sag_p2_TC1_cerv	*
II_t2_tse_sag_p2_TT1_thor	*
I t2_tse_stir_sag_p2_2C1	*
II t2_tse_stir_sag_p2_thor_2T1	*
I_t2_space_stir_iso_cerv_3C1	*
II_t2_space_stir_iso_thor_3T1	*
I_t1_tir_sag_PC2_cerv	*
II_t1_tir_sag_PT1_Thor	*
t2_tse_tra	*
I_pd_space_stir_iso_cerv_3C1	*
II_pd_space_stir_iso_thor_PrT1	*
I_pd_space_stir_iso_cerv_3C2_PF_on	*

pwal

t2_space_stir_sag_p3_iso_pwal	*
t2_space_stir_sag_p3_iso_få snit_56	*
t2_space_stir_sag_p3_iso_få snit thor	*
t1_tir_sag_Siemens	*
t1_tir_sag	*
t1_tir_sag_cerv_FP-1	*
t1_tir_sag_thor_FP-1	*
t2_space_dark-fluid_sag_p2_iso	*
t1_tir_sag_Siemens	*
t1_tir_sag_Siemens_TF	*
t1_tir_sag_Siemens_TF	*
t1_tir_sag_Siemens_TF	*
t1_tir_sag_TF	*
t1_tir_sag_TF	*
t1_tir_sag_TF	*
t1_tir_sag_TF	*
t2_space_stir_64snit_TF	*
t2_space_stir_64snit_TF	*
t2_space_stir_iso_Siemens_TF	*
t2_space_stir_1.5mm_p3_snit_56_4min	*
t2_space_stir_iso_Siemens_TF	*
t2_space_stir_iso_Siemens_TF	*
t2_space_stir_sag_p3_iso_Siemens	*
t2_space_stir_64snit_TF_thor	*
I_AASpine_Scout	*
II_AASpine_Scout	*
III_AASpine_Scout	*

- 1 -

Appendix 2: all results tables and figures.

Results

Table 1 showing data on study participants

MS Type	Number Participants	Female/ Male	Average Age (Yrs)	s.d	Average time from diagnosis (Yrs)	s.d (yrs)
CIS	3	3 / 0	54.7	3.21	2.00	5.859465
PPMS	1	1 / 0	65	-	3.00	-
RRMS	18	13 / 5	49	9.13	8.53	7.210798
SPMS	3	3 / 0	53.3	12.90	19.67	7.653975

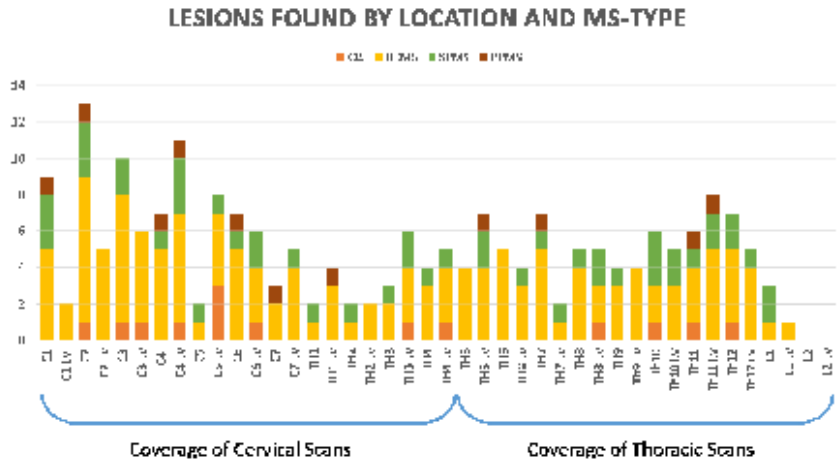

Figure 1 shows the number of lesions detected at consensus by location and MS type. The diagram also indicates which sequence a lesion found at a particular level was evaluated on for detection and CNR measurements.

Table 2 Number of lesions found for each MS subtype by consensus.

MS Type	Number Participants	Total number lesions)	Average number Lesions (S.D)
CIS	3	14	4.7 (2.5)
PPMS	1	11	11
RRMS	18	138	7.7 (4.8)
SPMS	3	47	15.7 (2.3)

Appendix 3: Images with and without correction for concomitant field effects.

Appendix 3: Images with and without correction for concomitant field effects, as tested with wrist images. The only difference between images is image correction. According to the paper, there are no differences in acquisition parameters. These figures and their descriptions are taken from Weaver et al 2018: Weavers PT, Tao S, Trzasko JD, Frigo LM, Shu Y, Frick MA, et al. B0 concomitant field compensation for MRI systems employing asymmetric transverse gradient coils. Magn Reson Med. 2018;79(3):1538-44.

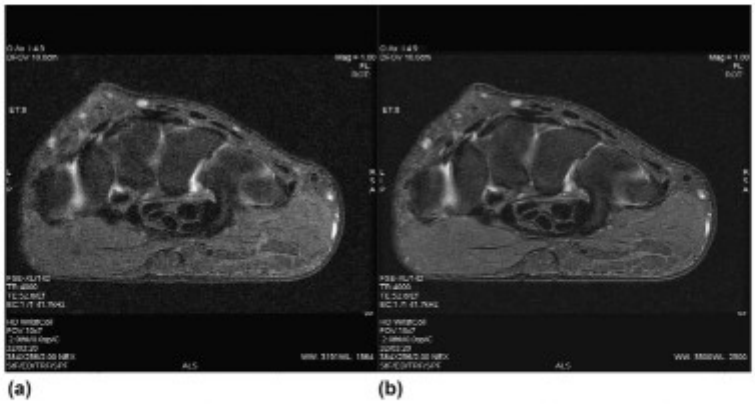

FIG. 4. Comparison wrist images without (a) and with (b) the B_0 concomitant field correction. Note the general decrease in signal, as well as overall blurring associated with imaging without compensation for the B_0 concomitant field. FSE, fast spin echo; TE, echo time; TR, repetition time.

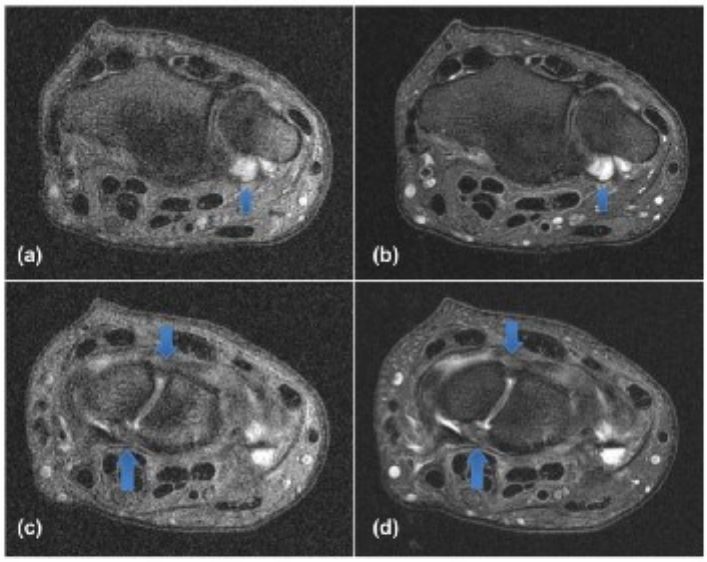

FIG. 5. Comparison wrist images without (a,c) and with (b,d) the B_0 concomitant field correction. Note the inability to differentiate between joint fluid and synovial thickening without compensation for the B_0 concomitant field. Also note the general decreased definition and conspicuity of the volar and dorsal components of the scapholunate ligament without compensation (c) for the B_0 concomitant field.

NTNU
Norwegian University of
Science and Technology

REGION ZEALAND
ZEALAND UNIVERSITY HOSPITAL