Comparison of MERGE and Axial T2-Weighted Fast Spin-Echo Sequences for Detection of Multiple Sclerosis Lesions in the Cervical Spinal Cord

Nancy Martin1
David Malfair2
Yinshan Zhao3
David Li1
Anthony Traboulsee3
Donna Lang4
Alexandra Talia Vertinsky1

OBJECTIVE. The purpose of our study was to compare axial multiple-echo recombined gradient echo (MERGE) with axial T2-weighted fast spin-echo (FSE) imaging for the detection of multiple sclerosis (MS) lesions in the cervical spinal cord on MRI.

MATERIALS AND METHODS. Twenty-nine cervical spine MRI studies of patients with MS lesions and 29 control cases were reviewed retrospectively. Two blinded neuroradiologists independently assessed randomized axial MERGE and axial T2-weighted FSE sequences from each study, documenting the location and number of cord lesions, the degree of confidence in calling each lesion, and the presence of artifacts. The reference standard was determined by an unblinded consensus review of all sequences performed for each case, with lesions considered present if detected on two or more sequences. Lesion detection rates and conspicuity, false-positive findings, and reader confidence and artifact scores were compared for the sequences, and interreader agreement was assessed.

RESULTS. Eighty-three lesions were assessed. The mean true-positive lesion detection rate was 87% (95% CI, 79–93%) with MERGE and 67% (60–75%) with T2-weighted FSE, with interreader positive agreement scores of 74% and 75%, respectively. A greater number of false-positive findings were seen with MERGE for both the MS and control cases. Average confidence and artifact scores were similar for both sequences. Subjectively, lesions were more conspicuous in 21 cases with MERGE and four cases with T2-weighted FSE and were equally conspicuous in four cases.

CONCLUSION. MERGE and T2-weighted FSE sequences are complementary. MERGE provided greater sensitivity for cord lesions whereas axial T2-weighted FSE provided improved lesion specificity. Further investigation is required to assess the clinical impact of MERGE in the diagnosis and management of MS.

Cervical spine MRI is important in the workup of demyelinating disease, particularly suspected multiple sclerosis (MS). Spinal cord abnormalities are seen on MRI in up to 83% of newly diagnosed MS patients [1] and support an atypical clinical presentation or equivocal brain MRI by ruling out alternative pathology and improving imaging sensitivity and specificity. Bot et al. [1] showed that the inclusion of a single spinal cord lesion in addition to brain imaging findings allowed 84.6% of newly diagnosed MS patients to meet the McDonald criteria for dissemination in space, compared with only 66.3% using brain imaging alone. White matter lesions detected in the brain by MRI are often nonspecific, with significant overlap between MS and many other CNS inflammatory (e.g., systemic lupus erythematosus, sarcoidosis, Sjögren syndrome, intermediate uveitis) and ischemic processes. The presence of spinal cord lesions, however, is more specific to MS, with the presence of a cord abnormality alone providing up to 92% accuracy in identifying MS among other neurologic disorders [2].

Spinal cord abnormalities seen on MRI were formally incorporated into the McDonald diagnostic criteria for MS in 2005 [3]. The current recommendations for spinal cord imaging include axial and sagittal T2-weighted, sagittal T1-weighted, and sagittal STIR or proton density sequences, with the option for additional contrast-enhanced T1-weighted imaging for suspicious lesions and 3D inversion recovery prepared T1 gradient-echo sequences [4]. Despite advances in cardiac gating and phased-array coils, spatial and contrast resolution and motion and pulsation artifact remain limiting factors in current protocol acquisitions.
A number of sequences have been studied in an attempt to improve spinal cord lesion detection. The FLAIR sequence, despite superior lesion detection in the supratentorial brain, does not perform well in the spinal cord because it is hampered by CSF pulsation artifact, poor contrast-to-noise ratio, and poor lesion detection [5]. Although STIR has shown improved lesion detection and conspicuity over T2-weighted fast spin-echo (FSE) and T2-weighted conventional spin-echo imaging [5–8], poor signal-to-noise ratio (SNR) [5, 8, 9], interpretative limitations due to flow and motion artifacts, and poor interobserver agreement [10] have also been described. Standard gradient-echo imaging has shown improved lesion detection over T2-weighted FSE when used with magnetization transfer [6] but has limited gray-white matter contrast [11]. More recently, T1-weighted inversion recovery has shown promise, with improved contrast between demyelinating lesions and normal cervical cord relative to T2-weighted FSE and STIR as well as improved lesion localization and delineation, particularly with phase-sensitive reconstruction [12], but has not yet been widely implemented.

There is a relatively new gradient-echo sequence that combines multiple bipolar gradient-echo formations using early echoes to increase SNR and later echoes to increase image contrast. This sequence is known by a variety of acronyms depending on the equipment vendor: MERGE (multiple-echo recombined gradient echo) for GE Healthcare, MEDIC (multiecho data image combination) for Siemens Healthcare, and MFFE (multiecho fast field echo) for Philips Healthcare. To date, this sequence has most commonly been used to replace the standard T2* gradient-echo sequence for assessment of disk pathology in the cervical spine [11], benefiting from its reduced pulsation artifact relative to T2-weighted FSE, excellent bone-CSF-soft tissue contrast [11, 13, 14], and option of 3D acquisition. MERGE, however, also shows excellent gray-white matter resolution (Fig. 1), making it ideal for assessment of intramedullary pathology [11, 13]. A small study comparing MEDIC with magnetization transfer to standard T2-weighted turbo spin-echo, STIR, and T1 spin-echo unenhanced and contrast-enhanced sequences has shown superior or visibility of intramedullary hemorrhage and mild cord edema as well as improved contrast between spinal cord gray matter, white matter, and edema on MEDIC [15]. To date, however, there have been no studies comparing MERGE to standard sequences in the MRI workup of MS. In this article, we compare the ability of MERGE and axial T2-weighted FSE to detect focal T2-hyperintense lesions in the cervical spinal cord.

Materials and Methods

Case Selection

The study was approved by the institutional clinical ethics research board. Our institutional database was searched from January 2005 to December 2008 for cervical spine MRI cases reported to have focal spinal cord lesions consistent with MS. Patients were included if they had focal T2-hyperintense lesions and if at least sagittal and axial T2-weighted FSE and axial MERGE sequences were performed as part of a standard clinical imaging protocol. Control cases were identified in a similar fashion by searching the institutional database for cases performed for indications other than MS and in whom the spinal cord was reported as normal. Only the 29 most recent control subjects were selected to match the number of MS cases.

MRI Parameters

The studies were acquired on one of two 1.5-T systems (Signa, GE Healthcare) with a phased-array head, neck, and spine coil using a standardized clinical protocol that included sagittal FSE T2-weighted (TR range/TE range, 3000–3500/102–110; echo-train length, 15–23; bandwidth, 16–31 kHz; matrix, 256–320 × 256; slice thickness, 3 mm; 0.5- to 1-mm skip; FOV, 20 cm; and number of signals acquired, 4), axial T2-weighted FSE (TR range/TE range, 2800–2500/102–110; echo-train length, 15–26; bandwidth, 16–31 kHz; matrix, 256 × 224 slice thickness, 3 mm; 1-mm skip; FOV, 20 cm; and number of signals acquired, 4), and axial MERGE (TR range, 950–1000; bandwidth, 31 kHz; matrix, 288 × 192; slice thickness, 3 mm; 1-mm skip; FOV, 20 cm; and number of signals acquired, 2). A variable number of additional sequences were performed in each case, including sagittal T1-weighted, sagittal STIR, and sagittal or axial proton density-weighted.

Image Evaluation

The MERGE and axial T2-weighted FSE sequences from both healthy and MS patients were randomized and presented separately on workstations for independent review by two neuroradiologists who were blinded to the clinical information. For each sequence, the reviewers documented the location of any cord lesions, assigned a confidence score in calling each lesion (from 1 indicating low confidence to 5 indicating high confidence), and assigned an overall sequence artifact score to each series of images from 1 indicating severe to 5 indicating none). After completion of the independent reviews, the reference standard of spinal cord lesions was determined by an unblinded consensus review of the MS cases by one of the neuroradiologists and a third radiologist with experience in neurologic MRI. Spinal cord lesions were considered present if visible on two or more of the available sequences in a given case. In review of each case during the consensus review, MERGE and T2-weighted FSE were also directly compared side-by-side for overall lesion conspicuity, and the false-positive lesions among the control subjects were individually examined for characterization. One of the MS cases was rejected during the consensus review because of severe patient motion artifact affecting all sequences and significantly limiting assessment.

Statistical Analysis

The number of lesions identified independent-ly by a given reader using MERGE and axial T2-weighted FSE sequences in isolation were compared with each other and to the reference lesion...
count using the Wilcoxon signed rank test. Sequence sensitivities for true-positive lesions were calculated, with 95% CIs obtained for each reading using the cluster bootstrapping method [16] and each subject as a sampling unit. The sensitivities of the two readings from the same reader were compared using the two-sided Mosteller exact test [17].

The mean confidence scores assigned to lesions by the same reader were compared using the z score, with MS and control cases analyzed separately. The standard error of the mean difference of these scores was obtained using cluster bootstrapping. We first compared the scores for all lesions called by a given reader, then for the true-positive and false-positive lesions separately. Because of small lesion counts, p values were not calculated for the false-positive lesions among the control groups.

The interreader agreement was measured by the percentage of positive agreement index [18]. This index reflects the chance that if one reader identified a site as a lesion, another reader will also identify the same site. Under perfect agreement, the probability is 1. Standard errors and CIs were obtained using the bootstrapping method. Sequence artifact scores assigned by the same reader were compared using the Wilcoxon signed rank test.

**Results**

The consensus review identified 83 reference-standard lesions among the MS cases. The total number of lesions (both true- and false-positive) identified during the independent case review were significantly higher with MERGE compared with T2-weighted FSE: 28% higher for reader 1 ($p = 0.01$) and 65% higher for reader 2 ($p = 0.0007$).

After comparison with the reference standard, the sensitivity for true-positive lesions, or the true-positive detection rate, was also higher using MERGE—significantly so for reader 2 ($p < 0.05$) but not for reader 1 ($p = 0.06$) (Table 1). The mean true-positive lesion detection rate for both readers was 87% (95% CI, 79–93%) for MERGE and 67% (60–75%) for T2-weighted FSE. Although total lesion counts and the true-positive detection rate varied significantly between the two readers using MERGE but not using T2-weighted FSE, lesion-by-lesion true-positive interreader percentage positive agreement was similar for both sequences: 74% for MERGE and 75% for T2-weighted FSE. False-positive lesion counts for the MS patients were significantly higher with MERGE relative to T2-weighted FSE for both readers ($p = 0.05$) (Table 1).

False-positive lesions were also called in the control cases using both sequences. Reader 1 called a total of eight false-positive lesions using MERGE compared with two using T2-weighted FSE, which could have mistakenly classified four cases and one case as potential MS, respectively. For reader 2, 16 false-positive lesions were called using MERGE and eight using T2-weighted FSE, with the potential for misclassifying six cases as possible MS for both sequences. Although false-positive counts in the control cases were higher using MERGE, the difference did not reach statistical significance ($p = 0.27$ for reader 1 and $p = 0.36$ for reader 2). The false-positive lesions seen with MERGE in the control cases were attributable to motion artifact (7/23 lesions), poor SNR (4/23), blurring of the gray matter (9/23), and misinterpretation of prominent gray matter signal and the central canal in the high cervical cord (3/23). Of note, susceptibility artifact did not impact cord signal and was not a source of false-positive findings with MERGE. As for T2-weighted FSE, false-positive lesions were related to CSF pulsation artifact (4/9), poor signal to noise (4/9), and blurring of the white matter (1/9).

The mean confidence scores assigned to all lesions are summarized in Table 2. Confidence scores ranged from 2 to 5, with a low confidence score of 2 being assigned only twice: once to a false-positive lesion within a control case and the other to a true-positive lesion—both cases using T2-weighted FSE. Except for reader 2, who had higher confidence scores for true-positive lesions with the MERGE sequence, mean confidence scores did not differ significantly between sequences. There was, however, suggestion of a trend toward progressively lower confidence scores.

---

**TABLE 1: Lesion Counts in Multiple Sclerosis Cases**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Reader 1</th>
<th>Reader 2</th>
<th>Reader 1</th>
<th>Reader 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>True-positive lesions</td>
<td>67</td>
<td>78</td>
<td>58</td>
<td>53</td>
</tr>
<tr>
<td>Reference true-positive lesions (%)</td>
<td>81</td>
<td>94</td>
<td>70</td>
<td>64</td>
</tr>
<tr>
<td>95% CI (%)</td>
<td>69–90</td>
<td>89–98</td>
<td>59–80</td>
<td>55–75</td>
</tr>
<tr>
<td>False-positive lesions</td>
<td>20</td>
<td>21</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Total lesions</td>
<td>87</td>
<td>99</td>
<td>68</td>
<td>60</td>
</tr>
</tbody>
</table>

*Note—MERGE = multiple-echo recombined gradient echo, FSE = fast spin-echo.

*Also referred to as “true-positive detection rate” or “sensitivity.”

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Reader 1</th>
<th>Reader 2</th>
<th>Reader 1</th>
<th>Reader 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean confidence score per lesion</td>
<td>4.56</td>
<td>4.68</td>
<td>0.39</td>
<td>4.71</td>
</tr>
<tr>
<td>Mean confidence score per true-positive lesion</td>
<td>4.67</td>
<td>4.72</td>
<td>0.70</td>
<td>4.87</td>
</tr>
<tr>
<td>Mean confidence score per false-positive lesion in MS cases</td>
<td>4.20</td>
<td>4.40</td>
<td>0.54</td>
<td>4.10</td>
</tr>
<tr>
<td>Mean confidence score per false-positive lesion in control cases</td>
<td>3.50</td>
<td>3.00</td>
<td>NA</td>
<td>3.59</td>
</tr>
</tbody>
</table>

*Note—Confidence was scored on 5-point scale from 1 (lowest) to 5 (highest), degree of confidence in calling each lesion. Because of insufficient counts, p values were not calculated for the control cases. MERGE = multiple-echo recombined gradient echo, FSE = fast spin-echo, MS = multiple sclerosis, NA = not applicable.

*Both true-positive and false-positive lesions called by a given reader.
Discussion

Evaluation of the spinal cord with MRI is challenging because of the small size of the spinal cord, CSF pulsation artifact, and susceptibility and ringing artifacts from the proximity of the bony spinal canal. Many sequences have been studied in an attempt to optimize spinal cord lesion detection. MERGE is a relatively new sequence with excellent gray-white matter contrast that has shown promise in identifying edema and hemorrhage within the cervical spinal cord [15], but its strengths and weaknesses for identifying MS lesions have not been characterized.

In our study, MERGE showed improved sensitivity for the detection of spinal cord lesions, with more true-positive lesions than T2-weighted FSE (Fig. 3). One reason for the higher true-positive lesion detection is the improved anatomic depiction of the normal H-shaped bright central gray matter on a background of low-signal white matter (Fig. 1), which permits improved distinction of pathologic high-signal foci. When the gray-white boundary is less well defined, pathologic signal change not completely in the periphery of the cord may be confused with normal central gray matter (Figs. 2A and 2B). The increased conspicuity of lesions on MERGE relative to T2-weighted FSE in 72% of cases also likely contributed to this increased sensitivity (Figs. 2C and 2D).

Although one might have expected the confidence scores in calling a lesion to reflect lesion conspicuity, our subjective assessment that cord lesions were more conspicuous on MERGE did not translate into higher confidence ratings. This is explained, at least in part, by the fact that confidence scores below 3 were almost never assigned; a lesion was unlikely to be called if confidence was low. In that light, improved lesion conspicuity with MERGE has not been reflected on a lesion-by-lesion basis but rather by improved sensitivity for detecting an increased number of lesions conspicuous enough to prospectively call with confidence.

At the expense of higher lesion detection with MERGE is a higher number of false-positive findings and the potential misclassification of normal control cases as MS cases. Some of these lesions may simply have not been shown well enough on the non-MERGE
sequences to be included in the reference standard. Because the sagittal T2-weighted FSE sequence was one of the main sequences used to confirm lesions present in the axial plane, bias may have been introduced to categorize lesions detected on T2-weighted FSE as true-positive lesions and those on MERGE as false-positive lesions. However, the presence of increased false-positive lesions on MERGE within the control cases indicates that overcalls do occur and will be a limitation of MERGE if viewed independently. Another possible reason for the higher false-positive rate with MERGE is the lack of familiarity with this sequence compared with standard T2-weighted FSE. Greater clinical experience with MERGE may translate into improved specificity and confidence in assessment of spinal cord lesions with MERGE over time.

Although confidence scores did not differ significantly between sequences, they did reveal an interesting trend of progressively lower confidence scores assigned in descending order of true-positive lesions, false-positive lesions among the MS cases, and false-positive lesions among the control cases. Lower confidence scores among the false-positive lesions most likely reflect simply a questionable imaging appearance, either due to decreased conspicuity of an individual lesion or suspicion that the finding represents artifact or even normal anatomy.

Although confidence scores were assigned on a per-lesion basis, this subjective assessment is influenced not only by their appearance but also by the presence or absence of lesions at other levels. The presence of multiple lesions will increase confidence in calling an additional lesion that individually may be less conspicuous. Thus, lesions with a questionable appearance, requiring correlation with other sequences, received lower confidence scores, whereas the presence or absence of additional cord lesions further stratified the false-positive lesions among MS and control cases, respectively. In practice, questionable lesions seen on one sequence would be correlated with additional sequences for confirmation or rejection, reducing the false-positive rate of both sequences.

Degradation of images by patient motion was the main source of artifacts in the MERGE images and a significant source of false-positive lesions called with MERGE in the control cases. Although MERGE shows excellent gray-white matter differentiation under ideal circumstances, when motion artifact is present, there is substantial blurring of the bright central gray matter. This motion artifact may be a reason for decreased specificity in some cases and may also explain the variability of the artifact scores assigned to MERGE. In comparison, the decreased variability of artifact scores assigned to T2-weighted FSE suggests the presence of CSF pulsation artifact—despite being present in more than one third of cases and accounting for one half of the control case false-positive lesions—has less impact on cord evaluation than motion does when using MERGE.

Our study has several limitations. First, our reference standard consisted of a review of all sequences performed rather than pathologic confirmation. The presence of both axial and sagittal T2-weighted FSE sequences, but only axial MERGE, introduces a bias against lesions seen only on the single axial MERGE sequence. Additionally, the retrospective review of axial sequences in isolation is artificial and limits our ability to translate the differences in sensitivity and specificity of MERGE and T2-weighted FSE into the setting of a complete imaging protocol and clinical practice.

Conclusion

This head-to-head comparison of axial MERGE and axial T2-weighted FSE sequences has shown these sequences to be complementary, each providing unique advantages in the identification of MS lesions. The superior gray-white matter differentiation within the spinal cord and the increased contrast and conspicuity of cord lesions with MERGE relative to T2-weighted FSE allow improved detection of intramedullary lesions but at the price of a higher false-positive rate. Thus, although axial T2-weighted FSE has greater specificity, MERGE shows greater sensitivity, suggesting the two sequences together have the potential to improve the workup of patients with suspected MS. Further investigations are warranted to assess the impact of including MERGE in current imaging protocols in meeting diagnostic criteria for MS and changing subsequent clinical management.

References

7. Campi A, Pontesilli S, Gerevini S, Scotti G, Com...

FOR YOUR INFORMATION

Unique customized medical search engine service from ARRS! ARRS GoldMiner® is a keyword- and concept-driven search engine that provides instant access to radiologic images published in peer-reviewed journals. For more information, visit http://goldminer.arrs.org.