

# Neuroradiologists Compared with Non-Neuroradiologists in the Detection of New Multiple Sclerosis Plaques

W. Wang, J. van Heerden, M.A. Tacey, and F. Gaillard

## ABSTRACT

**BACKGROUND AND PURPOSE:** Multiple sclerosis monitoring is based on the detection of new lesions on brain MR imaging. Outside of study populations, MS imaging studies are reported by radiologists with varying expertise. The aim of this study was to investigate the accuracy of MS reporting performed by neuroradiologists (someone who had spent at least 1 year in neuroradiology subspecialty training) versus non-neuroradiologists.

**MATERIALS AND METHODS:** Patients with  $\geq 2$  MS studies with 3T MR imaging that included a volumetric T2 FLAIR sequence performed between 2009 and 2011 inclusive were recruited into this study. The reports for these studies were analyzed for lesions detected, which were categorized as either progressed or stable. The results from a previous study using a semiautomated assistive software for lesion detection were used as the reference standard.

**RESULTS:** There were 5 neuroradiologists and 5 non-neuroradiologists who reported all studies. In total, 159 comparison pairs (ie, 318 studies) met the selection criteria. Of these, 96 (60.4%) were reported by a neuroradiologist. Neuroradiologists had higher sensitivity (82% versus 42%), higher negative predictive value (89% versus 64%), and lower false-negative rate (18% versus 58%) compared with non-neuroradiologists. Both groups had a 100% positive predictive value.

**CONCLUSIONS:** Neuroradiologists detect more new lesions than non-neuroradiologists in reading MR imaging for follow-up of MS. Assistive software that aids in the identification of new lesions has a beneficial effect for both neuroradiologists and non-neuroradiologists, though the effect is more profound in the non-neuroradiologist group.

**ABBREVIATIONS:** NNR = non-neuroradiologist; NR = neuroradiologist; VT = VisTarsier

Multiple sclerosis is the most common disease of the central nervous system in young patients, with a major impact on patients' lives.<sup>1</sup> Over the past decade, a number of disease-modifying drugs that are especially effective during early disease have been developed. Neurologists increasingly are aiming for zero disease progression and, in many instances, will alter management when progression is detected.<sup>2</sup> Because most demyelinating lesions are clinically occult, MR imaging has become the primary biomarker for disease progression. Both physical and cognitive disability have been shown to have a nonplateauing association with white matter demyelinating lesion burden as seen on T2-weighted and T2-weighted FLAIR sequences.<sup>2-6</sup> Detecting new

lesions can be an arduous task, particularly when there is a large number of pre-existing lesions.

Outside of study populations, MS MR imaging studies are reported by radiologists with varying expertise, ranging from general radiologists to fellowship-trained neuroradiologists. Although the accuracy of neuroradiologists (NRs) versus non-neuroradiologists (NNRs) has been examined in a number of settings, with varying results,<sup>7,8</sup> to our knowledge, no studies to date have investigated the efficacy of NR versus NNR reporting for MS.

We aimed to investigate the accuracy of MS reporting performed by NRs versus NNRs, with results from a previously published validated semiautomated assistive software platform (VisTarsier; VT) as a "gold standard." In this study, we hypothesized that nonspecialty reporters would perform at a slightly lower accuracy compared with subspecialty reporters.

## MATERIALS AND METHODS

### Patient Recruitment

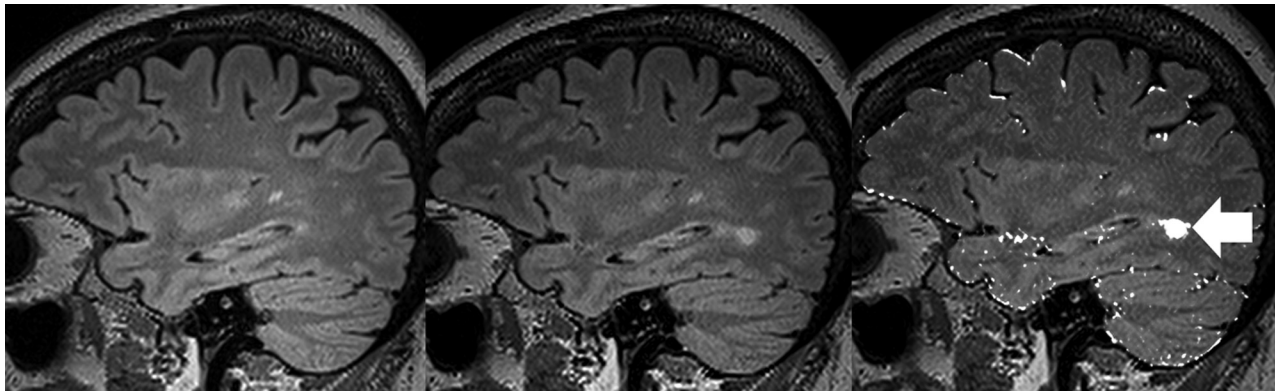
Institutional ethics board approval was obtained at the Royal Melbourne Hospital. The hospital PACS was queried for MR imaging

Received December 15, 2016; accepted after revision February 9, 2017.

From the Department of Radiology (W.W., F.G.) and Melbourne Epicentre (M.A.T.), the Royal Melbourne Hospital, Parkville, Victoria, Australia; and Perth Radiological Clinic (J.v.H.), Subiaco, Western Australia, Australia.

Please address correspondence to Dr Wayland Wang, Department of Radiology, the Royal Melbourne Hospital, Grattan St, Parkville 3050, Melbourne, Victoria, Australia; e-mail: wayland.wang@gmail.com

<http://dx.doi.org/10.3174/ajnr.A5185>



**FIG 1.** Example of VisTarsier output. From *left*: volumetric sagittal T2 FLAIR of previous study; volumetric sagittal T2 FLAIR of current study; and “New Lesions” map. NB: the usual output of the “New Lesions” map is colored in orange, whereas it is shown here colored *white*. The new lesion has been highlighted (*arrow*). There are slight misregistration errors at the edges of the brain, which can be dismissed as such by comparing the 2 studies on the left.

brain demyelination-protocol studies performed on a single 3T magnet (Tim Trio, 12-channel head coils; Siemens, Erlangen, Germany), between 2009 and 2011 inclusive, for patients who had  $\geq 2$  studies during that period. Eligibility criteria were the following: consecutive studies in patients with a confirmed diagnosis of MS (based on information provided on requisition forms), reported by a radiologist working at our institution, and availability of a diagnostic-quality MR imaging volumetric T2 FLAIR sequence (FOV, 250; 160 sections; section thickness, 0.98 mm; matrix,  $258 \times 258$ ; TR, 5000 ms; TE, 350 ms; TI, 1800 ms; 72 sel inversion recovery magnetic preparation).

### Data Analysis

A semiautomated software package (VT) has previously been validated and shown to detect an increased number of new lesions in patients with MS compared with conventional side-by-side comparison.<sup>9</sup> The development of the software processing tool and data processing procedure has been described in that earlier study. All of the patients included in this study have been reported previously.<sup>9</sup> The prior article described the development, implementation, and outcome of the semiautomated software package, whereas the current study reports the differences between NRs and NNRs in reporting MS studies by using the outcomes from the previous study as a “gold standard.”

All MS MR imaging studies were performed for routine clinical purposes and analyzed by a full-time academic radiologist at our institution (the Royal Melbourne Hospital) without the benefit of VT. These clinical radiology reports were obtained retrospectively from the institutional radiology information system archive. Each report was analyzed for new demyelinating lesions detected, which was recorded as a binary value of either “progression” (1 or more new lesions) or “stable” (no new lesions). For each reporting radiologist, we summarized the total number of reports and whether any new lesions were detected. This dataset had already been analyzed by 2 blinded neuroradiology-trained readers (J.v.H., F.G.) using VT and was used as a “gold standard” for lesion detection. An example of the VT output is displayed in Fig 1. Importantly, the reporting radiologists had the full diagnostic gamut of conventional imaging sequences available for review, excluding VT change maps, whereas the VT dataset used only the

change maps from VT (by using volumetric T2 FLAIR) in new lesion detection.<sup>9</sup>

Reporting radiologists were categorized as NR, defined as someone who had spent at least 1 year in a neuroradiology fellowship program, or NNR, defined as someone who has not had specific subspecialist training in neuroradiology.

### Statistics

There were 5 neuroradiologists and 5 non-neuroradiologists in our group. Patient demographic data were analyzed with a Student *t* test (unpaired) to test differences between the ages in the NR and NNR groups, and  $\chi^2$  tests were applied to test for differences between sex distributions in the 2 groups.

Sensitivity, specificity, positive predictive value, and negative predictive value were calculated for each group by using VT outcomes as the “gold standard.” A false-negative rate was also calculated for each group. These are presented as descriptive values.

As a further breakdown of NR versus NNR MS reporting, the variables associated with a false-negative reporting status were also assessed, with univariate analysis considering the effects of sex, age, number of lesions, and radiologist subgroup on the false-negative status. Variables identified with a *P* value  $< .20$  from univariate analysis were included in the multivariate model, and manual backwards stepwise regression techniques were used to identify those variables independently associated with false-negative status.

A 2-tailed *P* value  $< .05$  was considered to indicate statistical significance. STATA statistical analysis software (version 12.1) was used (StataCorp, College Station, Texas).

### RESULTS

In total, 159 comparison pairs (318 studies) met the above inclusion criteria. These pairs were drawn from 146 individual patients, of which 111 were women (76.0%) with a mean age of 44.9 years  $\pm$  10.7 (SD). Of the 159 pairs, 96 (60.4%) were reported by the 5 NRs, and 63 (39.6%) were reported by the 5 NNRs. Individual radiologist reporting numbers ranged from 7–43 for the NR group and 2–24 for the NNR group. There were no differences in patient demographics between the NR and NNR read groups

(mean age  $\pm$  SD, 44.3 years  $\pm$  9.7 versus 45.8 years  $\pm$  12.0;  $P = .415$ ; proportion of women, 76.4% versus 75.4%;  $P = .664$ ).

The VT-aided assessments reported 70 studies showing progression (ie, 1 or more new lesions). Of these, 39 were in the NR group and 31 in the NNR group. In comparison, 32 NR reports recorded progression, and 13 NNR reports recorded progression. The sensitivity, specificity, positive predictive value, and negative predictive value of NRs, NNRs, and the entire cohort in relation to new lesion detection are presented in Table 1. There was a higher level of sensitivity (82% versus 42%) and negative predictive value (89% versus 64%) in the NR group compared with the NNR group, with an associated lower level of false-negative reports (18% versus 58%) when treating VT-aided assessments as the “gold standard.”

An individual breakdown of true-positives and false-negatives

**Table 1: Sensitivity, specificity, positive predictive value, negative predictive value, and false-negative rate for neuroradiologists, non-neuroradiologists, and the combined cohort**

	NR	NNR	Combined
Sensitivity	82% (32/39)	42% (13/31)	64% (45/70)
Specificity	100% (57/57)	100% (32/32)	100% (89/89)
PPV	100% (32/32)	100% (13/13)	100% (45/45)
NPV	89% (57/64)	64% (32/50)	78% (89/114)
FN rate	18% (7/39)	58% (18/31)	36% (25/70)

**Note:**—FN indicates false-negative; NPV, negative predictive value; PPV, positive predictive value.

**Table 2: Reporting radiologist breakdown**

	Total	TP	FN <sup>a</sup>	FN Rate
NR1	10	4	0	0.0%
NR2	21	11	1	4.8%
NR3	15	7	1	6.7%
NR4	43	14	3	7.0%
NR5	7	3	2	29%
NNR1	21	9	4	19%
NNR2	5	0	1	20%
NNR3	24	13	6	25%
NNR4	11	6	5	45%
NNR5	2	0	2	100%

**Note:**—FN indicates false-negative; TP, true positive.

<sup>a</sup> A false-negative is defined as at least one new lesion detected using VT that was not detected in the clinical assessment of the study without the use of VT.

for each reporting radiologist is presented in Table 2. These data are also presented as a graph in Fig 2.

An assessment of the effect of age and sex on false-negative rates was not significant ( $P = .452$  for age and  $P = 1.00$  for sex) (Table 3). The NR group outperformed the NNR group, with 7/96 false-negatives compared with 18/63 ( $P < .001$ ). As expected, all false-negatives occurred when VT detected at least 1 new lesion, with the number of lesions also found to be higher for those patients with a false-negative status ( $P < .001$ ).

Multivariate analysis showed that the NNR group had a much higher false-negative rate compared with the NR group (OR, 7.88; 95% CI, 3.10–19.4;  $P < .001$ ), with having more than 1 lesion also being independently associated with false-negative status (OR, 11.94; 95% CI, 4.72–30.19;  $P < .001$ ).

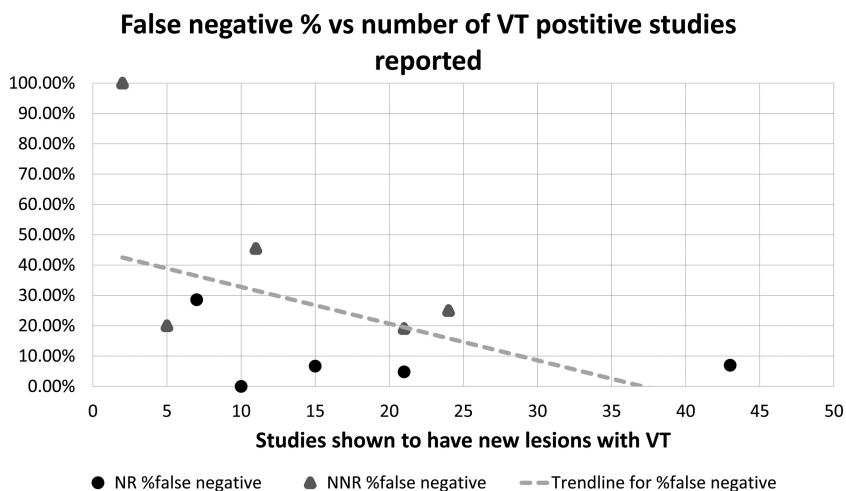
## DISCUSSION

The practice of MR imaging radiology is fairly heterogeneous depending on country, size of practice, and personal preference, with studies reported in some instances by general radiologists and at other times by subspecialty radiologists. Conventional MS follow-up study reading techniques have been shown to be subjective and dependent on the skill and consistency of the reader.<sup>10</sup> There have been many proposed solutions to facilitate time-efficient, reproducible, and accurate lesion detection, including semiautomated and fully automated techniques. Although many of these various proposed methods show promising results, the results are usually in small samples and generally have not entered routine clinical use.<sup>11</sup>

A previous study has shown that a semiautomated assistive platform is fast, robust, and detects many more new lesions on MR imaging of patients with MS than conventional side-by-side comparison.<sup>9</sup> The same software has also been shown to perform similarly when used by readers of different radiology experience, including readers with minimal radiology experience.<sup>12</sup> This new approach allowed us to use it as a new “gold standard” against which we can measure the performance of radiologists of various levels of subspecialty experience.

Other studies have shown no significant difference in NR versus NNR in stroke CT imaging,<sup>13</sup> significant discrepancy between NNR and second-opinion NR reporting,<sup>7</sup> and increased accuracy of interpretation in brain MR imaging by tertiary center NRs compared with external institutions.<sup>8</sup> Some of the aforementioned studies have showed the utility of second-opinion subspecialty reporting in neuroradiology.<sup>7,8</sup> However, in these studies, the original reports were over-read by an NR, whereas in our study, the 2 groups are independent of one another.

To our knowledge, no previous study has directly compared the performance of NNRs with subspecialty NRs in reporting a cohort of MR imaging for the same indication against a validated external “gold standard.”



**FIG 2.** False-negative percent versus all reported studies for all radiologists included in this study.

**Table 3: Univariate associations between variables and false-negative reporting status**

Variable	Correct	False-Negative	P Value
Total, no. (%)	134 (84.3)	25 (15.7)	N/A
Age, yr (Mean ± SD)	44.5 ± 10.6	46.2 ± 10.3	.452
Gender, no. (%)			1.000
Female	100 (74.6)	19 (76.0)	
Male	34 (25.4)	6 (24.0)	
Lesions found with VT	0 (0–1)	2 (1–3)	<.001
Lesions found with VT category, no. (%)			.031
≤2	111 (82.8)	16 (64.0)	
>2	23 (17.2)	9 (36.0)	
Lesions found with VT category, no. (%)			<.001
≤1	102 (76.1)	7 (28.0)	
>1	32 (23.9)	18 (72.0)	
Lesions found with VT category, no. (%)			<.001
≤0	89 (66.4)	0	
>0	45 (33.6)	25 (100.0)	
Group, no. (%)			<.001
NR	89 (66.4)	7 (28.0)	
NNR	45 (33.6)	18 (72.0)	
Reporter, no. (%)			.277
NR			
Reporter 1	10 (11.2)	0	
Reporter 2	40 (44.9)	3 (42.9)	
Reporter 3	14 (15.7)	1 (14.3)	
Reporter 4	5 (6.6)	2 (28.6)	
Reporter 5	20 (22.5)	1 (14.3)	
NNR			.123
Reporter 1	0	2 (11.1)	
Reporter 2	4 (8.9)	1 (5.6)	
Reporter 3	17 (37.8)	4 (22.2)	
Reporter 4	18 (40.0)	6 (33.3)	
Reporter 5	6 (13.3)	5 (27.8)	

Note:—N/A indicates not available.

The main finding of our study is that NRs performed better than NNRs in the conventional reading of MS studies in terms of better sensitivity, better negative predictive value, and lower false-negative rate. Specificity and positive predictive value were at an expected 100%, as there were no false-positives in our cohort. Among individual radiologists, the false-negative rate ranged from 0%–29% for NRs and 19%–100% for NNRs. The caveat is that the radiologist with a 100% false-negative rate had only reported 2 MS studies during the recruitment period.

In addition, we have also demonstrated that even when only grouping results into 2 or more lesions, versus 1 or 0 lesions, there is a much higher false-negative rate among NNRs compared with NRs (OR, 7.88) when taking into account other variables such as number of studies and patient demographics. This is potentially an important finding because, in some circumstances, a solitary asymptomatic lesion may not be sufficient for neurologists to consider treatment change.

There is a trend toward subspecialty reporting in North America. This is also the case in many other countries, though it is our direct experience that until all neuroradiology studies are reported by subspecialty-trained radiologists, general radiologists will choose to report studies that are perceived to require less subspecialty expertise. Follow-up MS studies are anecdotally of-

ten thought of as only requiring identification of new lesions, a task that is felt to be both cognitively simple and requiring less interpretation. Our findings suggest that this assumption is incorrect and that experience has a significant impact on the accuracy of detecting new lesions. As such, we believe this study not only supports the general move toward subspecialty reporting, but also suggests that there may not be such a thing as an “easy” study.

This study has a number of weaknesses. The largest and most difficult to control for is the fact that subspecialty-trained NRs also tend to report the greatest number of neuroradiology studies and would, in most cases, have done so more consistently over a longer period of time. We attempted to control for the overall number of neuroradiology studies reported by the included radiologist, though this proved to be impractical because of differences in years of experience, number of days worked during the study period, and variability in the type of studies some radiologists chose to report. As such, establishing the effect size of subspecialty training versus experience was not possible. Having said that, if looking at only the 2 most prolific reporters of MS studies from each group, each with more than 20 studies (having reported 64/96 of the studies from the NR group and 45/63 from the NNR group), the results are fairly similar to the entire cohort in that the false-negative rate in the NR group is much lower (6.3% versus 22%). Thus, we feel that the conclusion that NRs are more accurate than NNRs is broadly valid, even if the reason for this is not clear. It is important to note that we are not suggesting a simple direct causative link between subspecialty fellowship training and diagnostic accuracy, but rather using fellowship training as a pragmatic marker of a number of related factors that are likely to be contributory.

Another limitation of this study is that each study pair was reviewed by a single radiologist. Ideally, all 159 study pairs would have been read by each of the involved radiologists. However, this was beyond the scope of this study, which used pre-existing, retrospective data from an earlier study. In some ways, this is not a weakness, in that the clinical reports used are truly reflective of everyday practice rather than a somewhat artificial trial setting.

### Future Work

This work has focused on differences in traditional side-by-side comparison interpretation of MS MR imaging studies. It would be interesting, and is our plan, to evaluate how much individual benefit would be gained by radiologists of different subspecialty experience by using assistive software (VT). Another avenue of research is to plot the number of studies per year against performance and determine whether there is a plateau, which could inform a minimum of studies that should be read per year to maintain competence, as is the case in some countries for cardiac CT and CT colonoscopy.

### CONCLUSIONS

Neuroradiologists detect more new lesions than non-neuroradiologists in reading MR imaging for follow-up of MS. Assistive software that aids in the identification of new lesions has a beneficial effect for both neuroradiologists and non-neuroradiolo-

gists, though the effect is more profound in the non-neuroradiologist group.

## REFERENCES

1. Jones JL, Coles AJ. **New treatment strategies in multiple sclerosis.** *Exp Neurol* 2010;225:34–39 CrossRef Medline
2. Weiner HL. **The challenge of multiple sclerosis: how do we cure a chronic heterogeneous disease?** *Ann Neurol* 2009;65:239–48 CrossRef Medline
3. Mortazavi D, Kouzani AZ, Soltanian-Zadeh H. **Segmentation of multiple sclerosis lesions in MR images: a review.** *Neuroradiology* 2012;54:299–320 CrossRef Medline
4. Caramanos Z, Francis SJ, Narayanan S, et al. **Large, nonplateauing relationship between clinical disability and cerebral white matter lesion load in patients with multiple sclerosis.** *Arch Neurol* 2012;69:89–95 CrossRef Medline
5. Sormani MP, Rovaris M, Comi G, et al. **A reassessment of the plateauing relationship between T2 lesion load and disability in MS.** *Neurology* 2009;73:1538–42 CrossRef Medline
6. Fisniku LK, Brex PA, Altmann DR, et al. **Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis.** *Brain* 2008;131(Pt 3):808–17 CrossRef Medline
7. Briggs GM, Flynn PA, Worthington M, et al. **The role of specialist neuroradiology second opinion reporting: is there added value?** *Clin Radiol* 2008;63:791–95 CrossRef Medline
8. Zan E, Yousem DM, Carone M, et al. **Second-opinion consultations in neuroradiology.** *Radiology* 2010;255:135–41 CrossRef Medline
9. van Heerden J, Rawlinson D, Zhang AM, et al. **Improving multiple sclerosis plaque detection using a semiautomated assistive approach.** *AJNR Am J Neuroradiol* 2015;36:1465–71 CrossRef Medline
10. Abraham AG, Duncan DD, Gange SJ, et al. **Computer-aided assessment of diagnostic images for epidemiological research.** *BMC Med Res Methodol* 2009;9:74 CrossRef Medline
11. Vrenken H, Jenkinson M, Horsfield MA, et al. **Recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal studies of multiple sclerosis.** *J Neurol* 2013;260:2458–71 CrossRef Medline
12. Dahan A, Wang W, Tacey MA, et al. **Can medical professionals with minimal-to-no neuroradiology training monitor MS disease progression using semiautomated imaging software?** *Neurology* 2016;86(16 Supplement):P2.151
13. Jordan YJ, Jordan JE, Lightfoote JB, et al. **Quality outcomes of reinterpretation of brain CT studies by subspecialty experts in stroke imaging.** *AJR Am J Roentgenol* 2012;199:1365–70 CrossRef Medline