

A. Keshavan<sup>1</sup>, K. M. Leyden<sup>1</sup>, D. Peterson<sup>1</sup>, L. Bajus<sup>1</sup>, R. Markello<sup>1</sup>, M. Iv<sup>1,2</sup>, L. Tong<sup>1,2</sup>, Nina de Oliveira Soares Siebenborn<sup>3</sup>, C. Granziera<sup>4</sup>  
<sup>1</sup>Octave, Menlo Park, CA, USA; <sup>2</sup>Stanford University, Palo Alto, CA, USA; <sup>3</sup>University of Basel, Basel, Switzerland; <sup>4</sup>University Hospital Basel, Basel, Switzerland

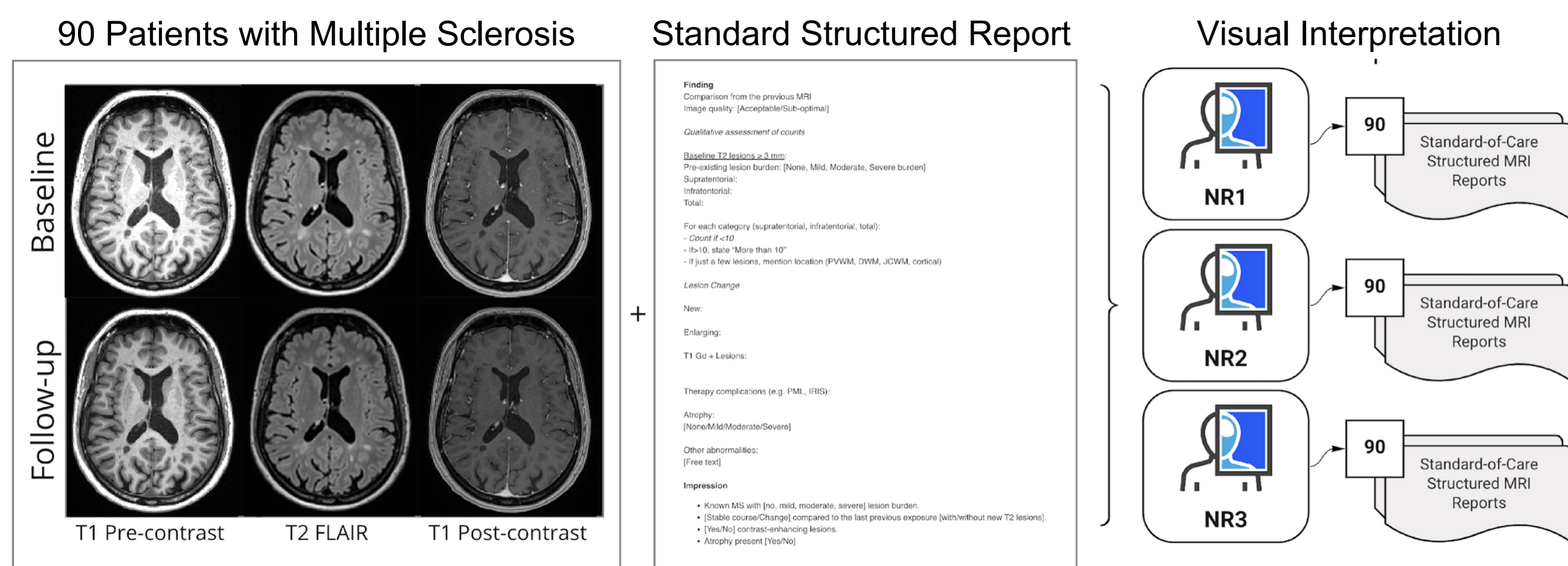
## INTRODUCTION

- New, enhancing, & enlarging lesions are key features neurologists use to determine whether patients with multiple sclerosis (MS) have disease activity that can impact treatment-related decisions.
- It is known that inter-rater variability exists among neuroradiologists' (NRs) reports, however evidence demonstrating this variability and its impact on clinical care is limited

## OBJECTIVE

This study aims to quantify the level of agreement between NRs' assessment of MS-specific standard of care reporting of brain magnetic resonance imaging (MRI) exams.

## METHODS



- 90 de-identified MS subjects with 2 (median 1 year apart) MRI exams from 2012 to 2019 were retrospectively enrolled from the University Hospital Basel.
- Enrolled subjects were randomly selected to represent varying levels of disease activity.
- Each MRI exam consisted of 3D pre- and post-gadolinium T1 and 3D T2 FLAIR sequences.
- Three fellowship-trained NRs (NR 1, NR 2, and NR 3) with up to 9 years of post-fellowship clinical experience visually interpreted the images for each subject and assessed report metrics shown in Table 2. The NRs were instructed to provide their interpretation in a standardized structured report template.
- Descriptive statistics were reported as frequency and percentage for categorical variables.
- Inter-rater reliability analysis with Fleiss' Kappa statistics was performed to evaluate the degree of consistency between NRs in their assessment of specific metrics in MRI reports. Percentage of agreement in each metric in MRI reports were also reported.
- The Fleiss' Kappa value was interpreted as  $\leq 0$  indicating no agreement, 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement.

## RESULTS

- Of the 90 MS subjects, 64.6% were female and the mean age was  $51.7 \pm 10.4$  years old, with an average EDSS score of  $3.3 \pm 1.7$ .
- Table 1 describes the distribution (number of reports, percentage of reports) of the assessment of report metrics in all 270 MRI reports, then separated into 90 MRI reports by each NR.
- Table 2 describes the inter-rater reliability analysis of report metrics between the 3 NRs.

## RESULTS

Table 1. Distribution of NR assessments across report metrics (Count, %)

Characteristic	Overall (N = 270)	NR 1 (N = 90)	NR 2 (N = 90)	NR 3 (N = 90)
<b>Image Quality</b>				
Acceptable	214 (80%)	59 (66%)	83 (92%)	72 (83%)
Suboptimal	53 (20%)	31 (34%)	7 (8%)	15 (17%)
<b>Preexisting Lesion Burden</b>				
None	3 (1%)	--	1 (1%)	2 (2%)
Mild	102 (38%)	27 (30%)	54 (60%)	21 (23%)
Mild-Moderate	2 (<1%)	--	--	2 (2%)
Moderate	99 (37%)	47 (52%)	19 (21%)	33 (37%)
Moderate-Severe	5 (2%)	--	--	5 (6%)
Severe	58 (21%)	16 (18%)	16 (18%)	26 (29%)
Significant	1 (<1%)	--	--	1 (1%)
<b>Brain Atrophy</b>				
None	109 (41%)	1 (1%)	54 (61%)	54 (61%)
Mild	118 (44%)	71 (79%)	25 (28%)	22 (25%)
Mild-Moderate	1 (<1%)	--	--	1 (1%)
Moderate	36 (13%)	16 (18%)	10 (11%)	10 (11%)
Moderate-Severe	2 (<1%)	--	--	2 (2%)
Severe	2 (<1%)	2 (2%)	--	--
<b>Description of Lesion Counts</b>				
<b>Supratentorial Lesions</b>				
1 to 10	23 (9%)	11 (12%)	2 (2%)	10 (11%)
11 or more	246 (91%)	79 (88%)	87 (98%)	80 (89%)
<b>Infratentorial Lesions</b>				
0	85 (32%)	33 (37%)	29 (33%)	23 (26%)
1 to 10	159 (60%)	52 (58%)	46 (52%)	61 (69%)
11 or more	23 (9%)	5 (6%)	13 (15%)	5 (6%)
<b>Description of Lesion Change</b>				
New Lesions	51 (19%)	25 (29%)	13 (15%)	13 (14%)
Enlarging Lesions	20 (9%)	15 (24%)	1 (1%)	4 (5%)
Enhancing Lesions	10 (4%)	4 (4%)	3 (3%)	3 (3%)
<b>Count of New Lesions</b>				
0	212 (81%)	60 (72%)	75 (85%)	77 (86%)
1	35 (13%)	20 (24%)	9 (10%)	6 (7%)
2	7 (3%)	3 (4%)	1 (1%)	3 (3%)
4	5 (2%)	--	3 (3%)	2 (2%)
5	2 (<1%)	--	--	2 (2%)
<b>Count of Enlarging Lesions</b>				
0	216 (93%)	48 (79%)	87 (99%)	81 (96%)
1	12 (5%)	10 (16%)	1 (1%)	1 (1%)
2	4 (2%)	3 (5%)	--	1 (1%)
5	1 (<1%)	--	--	1 (1%)
<b>Count of Enhancing Lesions</b>				
0	258 (98%)	86 (96%)	86 (97%)	86 (97%)
1	4 (2%)	--	2 (2%)	2 (2%)
2	2 (<1%)	--	1 (1%)	1 (1%)

Table 2. Results of inter-rater reliability analysis across report metrics

	Percentage of Agreement (%)	Fleiss Kappa Coefficient	Interpretation
Brain Atrophy	11%	0.02	None-slight
Presence of Enlarging Lesions	50%	0.03	None-slight
Enlarging Lesion Count	79%	0.06	None-slight
Image Quality	59%	0.18	None-slight
New Lesion Count	68%	0.25	Fair
Infratentorial Lesion Count	26%	0.26	Fair
Supratentorial Lesion Count	86%	0.32	Fair
Presence of New Lesions	67%	0.34	Fair
Preexisting Lesion Burden	37%	0.35	Fair
Enhancing Lesion Count	99%	0.50	Moderate
Presence of Enhancing Lesions	96%	0.62	Substantial

## CONCLUSIONS

- Neuroradiologists were **most consistent** when reporting the **presence and count of enhancing lesions**.
- However, **all other metrics** of disease activity and progression, such as **new and enlarging lesions**, varied from **none-slight to fair agreement**.
- The varying agreement highlights the need to provide Neuroradiologists with tools to reduce variability in detecting clinically relevant features that impact patient care.
- Future studies will examine the impact of quantitative segmentation data such as lesion detection and atrophy measures on the content and intra- & inter-rater reliability of MRI brain reports for patients with MS.

