Semi-automatic detection of changes in T2 lesion load in multiple sclerosis in comparison to visual analysis: increased diagnostic yield?

Poster No.: C-3044
Congress: ECR 2017
Type: Scientific Exhibit
Authors: K. Schwerdtner¹, M. Domin¹, H.-J. Huppertz², M. Grothe¹, E. Rathmann¹, S. Langner¹; ¹Greifswald/DE, ²Zurich/CH
Keywords: Neuroradiology brain, Computer applications, MR, Computer Applications-Detection, diagnosis, Comparative studies
DOI: 10.1594/ecr2017/C-3044

Any information contained in this pdf file is automatically generated from digital material submitted to EPOS by third parties in the form of scientific presentations. References to any names, marks, products, or services of third parties or hypertext links to third-party sites or information are provided solely as a convenience to you and do not in any way constitute or imply ECR's endorsement, sponsorship or recommendation of the third party, information, product or service. ECR is not responsible for the content of these pages and does not make any representations regarding the content or accuracy of material in this file.

As per copyright regulations, any unauthorised use of the material or parts thereof as well as commercial reproduction or multiple distribution by any traditional or electronically based reproduction/publication method ist strictly prohibited.

You agree to defend, indemnify, and hold ECR harmless from and against any and all claims, damages, costs, and expenses, including attorneys' fees, arising from or related to your use of these pages.

Please note: Links to movies, ppt slideshows and any other multimedia files are not available in the pdf version of presentations.

www.myESR.org
Aims and objectives

Background - Epidemiology MS

Multiple sclerosis (MS) is the most common inflammatory-demyelinating disease of the central nervous system. With a prevalence greater than 100 per 100,000, it affects over 2 million people worldwide. [1, 2]

It is the most frequent cause of non-traumatic neurological disability in young and middle-aged adults [3] and there has been an increasing incidence over the last decades. [4]

Two third of the patients are women and the typical age of manifestation is between 20 and 40 years. [1]

The symptoms are characterised by great variety and diversity and in fact there are no neurological symptoms which cannot be related to MS. [1]

Background - Diagnosis of MS

MS is a clinical diagnosis, which needs detailed medical history, clinical examination and supportive paraclinical investigations like MRI, CSF or evoked potentials. Symptoms have to be disseminated in time and space and must not be caused by another disease. [5]

Demyelinating white matter lesions are a marker for disease activity, severity of symptoms and the progression of disability and hence became a useful parameter for diagnostics and monitoring of MS. [6, 7]

MRI is the most sensitive technique for detection of MS-plaque [8]. Hence it became important for the detection of changes in lesion load in T2-weighted images (T2ll) in order to monitor both disease activity and effects of therapy.

In clinical routine T2ll is assessed by visual evaluation of the MR images which is a time consuming task and prone to user bias.
Therefore it is desirable to support the assessment of changes in T2ll (cT2ll) with computer-based tools.

Until today plenty of algorithms have been developed for lesion detection and volumetry, [9, 10, 11, 12, 13, 14, 15] but no gold standard has been defined yet.

Furthermore, the majority of these techniques have only been evaluated in small trials.

**Aim of the Study**

The aim of the present study was to evaluate diagnostic yield and accuracy of semi-automatic detection of changes in T2ll (sadcT2ll) changes in comparison to visual analysis.
Methods and materials

Material & Methods - Patients

The study was performed in accord with the Helsinki Declaration of 1975 and approved by the local ethics committee. Written informed consent was waived.

We conducted a retrospective study at our institution on MRI data achieved between September 1\textsuperscript{st}, 2009 and April 30\textsuperscript{th}, 2015.

Every patient who received a baseline MRI and at least one follow-up study under suspicion or for control of a chronic demyelinating central nervous disease was included.

Material & Methods - MR imaging

All MR examinations were performed at 3T (Magnetom Verio, Siemens, Erlangen; Syngo Vers. VB 17A) using 32-channel head coils for signal detection.

Imaging protocol consisted of at least:

1. sagittal 3D T1-weighted MPR dataset with 0.9 mm isotropic spatial resolution (TE 2.58 ms, TR 1900 ms, TI 900 ms, flip angle 9°, matrix 246 x 256, FOV 220, TA 4:18 min)
2. axial T2-weighted TSE dataset with 0.6 x 0.5 x 3 mm spatial resolution (TE 103 ms, TR 6100 ms, TI 1900 ms, flip angle 150°, matrix 358 x 448, FOV 220, TA 4:12 min)

T1w datasets were used for spatial and signal intensity normalization.

T2w datasets were used for the assessment of T2ll.

For each patient the recent T2w dataset was compared with the T2w dataset of the immediate preceding examination.
Material & Methods - Image Processing

For this study an algorithm, which was introduced by Huppertz in 2011 [16], was modified.

For all patients and all available T2w datasets the following steps were performed:

1. Conversion of DICOM images to NIFTI format using dcm2nii [17]
2. The most recent T1w image was spatially normalized to Montreal Neurological Institute (MNI) space and segmented in gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). The tissue classes were corrected for signal intensity inhomogeneities.
3. Correction of the T2w images for signal intensity inhomogeneities.
4. Coregistration of T1w and T2w images in native space.
5. Normalization of the coregistered and bias corrected T2w image to MNI space.
6. For signal intensity normalization of the T2w dataset, the previously obtained results of GM and WM segmentation (step 2) were binarized and then used as masks to determine average intensities of GM and WM voxels in the T1 image and in the T2 image, respectively.
7. An intensity average of the whole brain was calculated as the mean of both tissue class intensities.
8. After that the global brain intensity was set to an arbitrary value of 1000 and the resulting modulating factor applied to each voxel to achieve the intensity normalization of the brain and make MRI scans comparable to each other.
9. Calculating a difference image (SI) between baseline and follow-up T2w image after coregistration and intensity normalization of both images using SPM12-based algorithm allowed the visual analysis of cT2Ii.

Material & Methods - Quality Check

To ensure high quality of the used MR images, the resulting images and its intermediate steps after every step the images were checked and excluded in case of lacking quality.

Material & Methods - Determination of changes in T2 Lesion Load

Changes in T2 lesion load were determined by evaluation of the subtractions images (SI) with MRIcro (Version 1.37 build 4).

Each SI was evaluated using standardized w/l-settings and standardized zoom.
Each change in SI clearly determined in three dimensions was counted as new lesion. Exceptions were made regarding changes which were clearly located within a previously existing lesion, and hence not counted as a new lesion (i.e. change within lesion).

Written radiological reports were used for assessment of visual evaluation of cT2ll.

**Material & Methods - Comparison between sadcT2ll and radiological findings**

Radiological assessment of changes in T2ll (racT2ll) and sadcT2ll were compared to each other in order to categorize the results as true positive (TP), false positive (FP), true negative (TN) and false negative (FN):

If both methods had the same results, they were both counted as TP - in case of change in T2ll - or as TN in case of no change, respectively.

If the results of both methods differed from each other, T2w images were reviewed by an experienced neuroradiologist and the results were categorized by arbitration for each method:

- **TP**: correctly assessed number of cT2ll - in case a change took place
- **FP**: falsely assessed number of cT2ll - either too many lesion changes found or detected change in SI without cT2ll
- **TN**: correctly assessed cT2ll - in case no change took place
- **FN**: missed lesion changes - either less or no detected changes in case a change took place

**Material & Methods - Statistics**

For statistical evaluation of the results, the assessed data were used to calculate statistical parameters: mean, range, prevalence for the T2 lesions.

Sensitivity, specificity, positive prediction value (PPV), negative prediction value (NPV), false positive rate (FPR), false negative rate (FNR), positive (LH+) and negative Likelihood Ratio (LH-) were calculated for the results regarding the quantitative assessment on cT2ll on the one hand and for the results disregarding quantitatively correct assessment on the other.

Calculations were performed with Microsoft Office Excel 2013.
Fig. 2: Overview of the image processing steps of the modified algorithm. The original algorithm was first introduced by Huppertz in 2011. [16]

© Institute for Diagnostic Radiology and Neuroradiology, University Medicine Greifswald - Greifswald/DE
Results

Results - Patients

During the study period, 357 patients with clinically proven or suspected chronic demyelinating central nervous disease underwent baseline MRI and at least one follow-up study.

- 77 patients were excluded due to MR imaging protocol violations.
- 9 patients were excluded due to incomplete segmentation process.
- In 16 patients the diagnosis of a chronic demyelinating central nervous disease could not be confirmed.

255 patients were included in the final evaluation. 85 were male (33.3%) and 170 female (66.7%), the mean age was 45.1 years (range = 18.0 - 77.8).

Results - MR imaging

255 patients had 804 MRI examinations (median = 3; mean = 3.15; range = 1 - 8).

From these 804 examinations 550 subtraction datasets could be generated (median = 2; mean = 2.16; range= 1 - 7).

19 subtraction datasets had to be excluded due to severe imaging artifacts and in 4 patients the written report was unavailable.

Final analysis comprised 531 subtraction datasets.

Results - Determination of changes in T2 Lesion Load and comparison values

sadcT2ll found a change 164 of 531 SI (30.89%), while radiological assessment found a change in 193 of 531 SI (36.35%).

Results of arbitration
After establishing the standard with arbitration as mentioned above cT2ll were found in 192 SI (36.16%).

The following results of a detected change regardless to its correct quantification:

- In 163 cases the result of sadcT2ll was TP compared to 178 cases of racT2ll, respectively.
- Results of sadcT2ll were FP in 1 cases, racT2ll in 8 cases, respectively.
- TN results occurred 338 times in sadcT2ll and 331 times in racT2ll, respectively.
- In 29 cases sadcT2ll was FN and in 14 cases racT2ll, respectively.

<table>
<thead>
<tr>
<th></th>
<th>true positive</th>
<th>false positive</th>
<th>true negative</th>
<th>false negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>sadcT2ll</td>
<td>163</td>
<td>1</td>
<td>338</td>
<td>29</td>
</tr>
<tr>
<td>racT2ll</td>
<td>178</td>
<td>8</td>
<td>331</td>
<td>14</td>
</tr>
</tbody>
</table>

*Table 1: Results for detection of cT2ll regardless of correct quantification*

Regarding the correct assessment of the number of cT2ll there were following results:

- In 133 cases the result of sadcT2ll was TP compared to 150 cases of racT2ll, respectively.
- Results of sadcT2ll were FP in 2 cases, racT2ll in 15 cases, respectively.
- TN results occurred 338 times in sadcT2ll and 331 times in racT2ll, respectively.
- In 58 cases sadcT2ll was FN and in 35 cases racT2ll, respectively.

<table>
<thead>
<tr>
<th></th>
<th>true positive</th>
<th>false positive</th>
<th>true negative</th>
<th>false negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>sadcT2ll</td>
<td>133</td>
<td>2</td>
<td>338</td>
<td>58</td>
</tr>
<tr>
<td>racT2ll</td>
<td>150</td>
<td>15</td>
<td>331</td>
<td>35</td>
</tr>
</tbody>
</table>

*Table 2: Results for quantitative detection of cT2ll*

**Results - Statistics**

The following table shows the results of both methods to detect changes of T2 lesion load regardless of its correct quantification.
Table 3

<table>
<thead>
<tr>
<th></th>
<th>sadcT2ll</th>
<th>racT2ll</th>
</tr>
</thead>
<tbody>
<tr>
<td>sensitivity</td>
<td>84.9%</td>
<td>92.7%</td>
</tr>
<tr>
<td>specificity</td>
<td>99.7%</td>
<td>97.6%</td>
</tr>
<tr>
<td>PPV</td>
<td>99.4%</td>
<td>95.7%</td>
</tr>
<tr>
<td>NPV</td>
<td>92.1%</td>
<td>95.9%</td>
</tr>
<tr>
<td>FPR</td>
<td>0.3%</td>
<td>2.4%</td>
</tr>
<tr>
<td>FNR</td>
<td>15.1%</td>
<td>7.3%</td>
</tr>
<tr>
<td>LH+</td>
<td>287.80</td>
<td>39.29</td>
</tr>
<tr>
<td>LH-</td>
<td>0.15</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Table 4

The following table shows the results of both methods to detect quantitatively correct changes of T2 lesion load:

<table>
<thead>
<tr>
<th></th>
<th>sadcT2ll</th>
<th>racT2ll</th>
</tr>
</thead>
<tbody>
<tr>
<td>sensitivity</td>
<td>69.6%</td>
<td>81.1%</td>
</tr>
<tr>
<td>specificity</td>
<td>99.4%</td>
<td>95.7%</td>
</tr>
<tr>
<td>PPV</td>
<td>98.5%</td>
<td>90.9%</td>
</tr>
<tr>
<td>NPV</td>
<td>85.4%</td>
<td>90.4%</td>
</tr>
<tr>
<td>FPR</td>
<td>0.6%</td>
<td>4.3%</td>
</tr>
<tr>
<td>FNR</td>
<td>30.4%</td>
<td>18.9%</td>
</tr>
<tr>
<td>LH+</td>
<td>118.38</td>
<td>18.70</td>
</tr>
<tr>
<td>LH-</td>
<td>0.31</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Fig. 3: Flowchart of patients' assessment in the study.
A)

Fig. 4: Patient with new lesion

© Institute for Diagnostic Radiology and Neuroradiology, University Medicine Greifswald - Greifswald/DE
Fig. 5: Changes within a previous lesion (cwl)

© Institute for Diagnostic Radiology and Neuroradiology, University Medicine Greifswald - Greifswald/DE
Fig. 6: Artefacts

© Institute for Diagnostic Radiology and Neuroradiology, University Medicine Greifswald - Greifswald/DE
Conclusion

Discussion - Main Findings

Visual Assessment (sensitivity of 81.1% for racT2ll) is time consuming and error-prone. Therefore, computer aided lesion detection is desirable.

Disregarding the quantitatively correct assessment of changes:

- sensitivity sadcT2ll vs racT2ll 84.9% vs 92.7%
- specificity sadcT2ll vs racT2ll 99.7% vs 97.6%

Regarding the correct assessment of cT2ll:

- sadcT2ll lacks of accuracy (i.e. sensitivity of 69.6% vs 81.1% of racT2ll)
- sadcT2ll with high specificity (i.e. 99.4% vs 95.7% of racT2ll)

The findings demonstrate, that the diagnostic performance of the method for sadcT2ll presented here is inferior to radiological assessment of cT2ll.

Discussion - Limitations

With the presented algorithm, semi-automatic detected changes in T2ll still have to be evaluated by visual inspection making this approach prone to user bias.

- No threshold is defined to differentiate between a new lesion, new lesions with just subtle changes in T2-SI, changes within a previously existing lesion or artefacts.
- New lesions with inconsistent signals hinder fully automated assessment.

Detected and assessed changes of T2-SI within lesions ("change within lesion") are not part of the current clinical classification for the diagnostics of Multiple Sclerosis - and they are not included in radiological reports.

Conclusion

The method of sadcT2ll presented is this study is inferior to standard visual assessment by a radiologist.
Especially in assessing quantitatively correct lesion changes sadcT2ll lacks accuracy and does not improve diagnostic yield.

Nevertheless it is a fast and robust technique and could serve as complementary tool in daily clinical routine.

It may be beneficial in patients with high T2ll or with subtle lesion changes.
References


