**Supplementary Material**

**A Retrospective Belgian Multi-Center MRI Biomarker Study in Alzheimer’s Disease (REMEMBER)**

Ellis Niemantsverdriet, Annemie Ribbens, Christine Bastin, Florence Benoit, Bruno Bergmans, Jean-Christophe Bier, Roxanne Bladt, Lene Claes, Peter Paul De Deyn, Olivier Deryck, Bernard Hanseeuw, Adrian Ivanoiu, Jean-Claude Lemper, Eric Mormont, Gaëtane Picard, Eric Salmon, Kurt Segers, Anne Sieben, Dirk Smeets, Hanne Struyfs, Evert Thiery, Jos Tournoy, Eric Triau, Anne-Marie Vanbinst, Jan Versijpt, Maria Bjerke, Sebastiaan Engelborghs

**Validation Process of MSmetrix**

MSmetrix has been specifically designed to measure atrophy in patients with multiple sclerosis (MS) and has been validated in these patients [1-5]. Within this validation process test-retest scans were obtained in 18 MS patients, which were scanned on two Philips Healthcare systems (Philips, Best, The Netherlands) on the same day; 1.5Tesla (T, Intera) and 3T (Achieva) [2]. This study showed that MSmetrix was more robust compared to SIENA (FSL, http://www.fmrib.ox.ac.uk/fsl), as the median percentage error of whole brain (WB) volume measurements between 1.5T and 3T scanners was 0.35%, as compared to 2.99% for SIENA.

Next, test-retest scans were obtained in 10 MS patients, which were scanned twice per scanner at three different scanners (General Electric (GE) Medical Systems Discovery MR750w, n=8; SIEMENS Skyra, n=8; Philips Medical Systems Achieva, n=7) [1]. This study demonstrated a small measurement error across the three 3T scanners with a median value of 0.13% (MSmetrix) and 0.17% (SIENA), which was not significantly different between the two different segmentation methods. The daily physiological processes were evaluated in three subjects scanned two times on 20 different days within a 31-day period, and a significantly smaller overall error for WB atrophy was detected when measures were analyzed by MSmetrix (0.19%) compared to SIENA (0.31%). In addition, no significant differences were observed between MSmetrix and SIENA with regard to the confidence interval (CI) of WB atrophy for 6-month intervals compared to the one-year interval.

The usage of T1 only versus T1 over FLAIR with MSmetrix was investigated in 33 MS patients. The difference between the two image protocols, and thus analyzing techniques, was 5.74 mL (median) with a Pearson R of 1.00 (unpublished data).

Moreover, studies have shown it is robust for different scanners without parameter tuning and provides accurate segmentations with a good reproducibility [3,4]. The accuracy was evaluated by comparing the output from MSmetrix-cross with Lesion Segmentation Tool (LST) [6] and Lesion-Topology-preserving Anatomical Segmentation (TOADS) [7], two expert reference segmentations, in 20 MS patients. In here, spatial overlap had a mean ±standard deviation (SD) of 0.67 ±0.11 and an interclass correlation coefficient equals 0.8, indicating a good volumetric agreement between MSmetrix and expert labelling [4]. Next, MSmetrix-long was evaluated by LST, and the median Dice score was 0.63 with a Pearson correlation coefficient equals to 0.96 [3]. Again a good agreement, and thus, MSmetrix-long is able to accurately and reproducibly measure new, enlarging, disappearing, shrinking, and static volumes in MS populations.

When MSmetrix, FreeSurfer, SIENA, and Statistical Parametric Mapping (SPM) were compared, differences in atrophy measurements were larger compared to typical atrophy rates observed in MS, even at WB level [5]. MSmetrix-cross behaved similar to SPM and MSmetrix-long to SIENA, both in terms of mean volume difference as well as proportional error.

In conclusion, MSmetrix is a robust method to analyze different MRI scans with different protocols and/or acquired at different scanner types.

**REFERENCES**

[1] Smeets D, Ribbens A, Sima DM, Cambron M, Horakova D, Jain S, Maertens A, Van Vlierberghe E, Terzopoulos V, Van Binst AM, Vaneckova M, Krasensky J, Uher T, Seidl Z, De Keyser J, Nagels G, De Mey J, Havrdova E, Van Hecke W (2016) Reliable measurements of brain atrophy in individual patients with multiple sclerosis. *Brain Behav* **6**, e00518.

[2] Lysandropoulos AP, Absil J, Metens T, Mavroudakis N, Guisset F, Van Vlierberghe E, Smeets D, David P, Maertens A, Van Hecke W (2016) Quantifying brain volumes for Multiple Sclerosis patients follow-up in clinical practice - comparison of 1.5 and 3 Tesla magnetic resonance imaging. *Brain Behav* **6**, e00422.

[3] Jain S, Ribbens A, Sima DM, Cambron M, De Keyser J, Wang C, Barnett MH, Van Huffel S, Maes F, Smeets D (2016) Two time point MS lesion segmentation in brain MRI: an expectation-maximization framework. *Front Neurosci* **10**, 576.

[4] Jain S, Sima DM, Ribbens A, Cambron M, Maertens A, Van Hecke W, De Mey J, Barkhof F, Steenwijk MD, Daams M, Maes F, Van Huffel S, Vrenken H, Smeets D (2015) Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *Neuroimage Clin* **8**, 367-375.

[5] Steenwijk MD, Amiri H, Schoonheim MM, de Sitter A, Barkhof F, Pouwels PJW, Vrenken H (2017) Agreement of MSmetrix with established methods for measuring cross-sectional and longitudinal brain atrophy. *Neuroimage Clin* **15**, 843-853.

[6] Schmidt P, Gaser C, Arsic M, Buck D, Forschler A, Berthele A, Hoshi M, Ilg R, Schmid VJ, Zimmer C, Hemmer B, Muhlau M (2012) An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *Neuroimage* **59**, 3774-3783.

[7] Shiee N, Bazin PL, Ozturk A, Reich DS, Calabresi PA, Pham DL (2010) A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *Neuroimage* **49**, 1524-1535.

**Supplementary Table 1.** Overview of the REMEMBER study population based on quality control (QC) of obtained volumes by MSmetrix.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Controls** | **SCD** | **MCI** | **AD dementia** | **Total** |
| Center 1 | 45 | 43 | 139 | 96 | 323 |
| Center 2 |  | 2 | 13 | 15 | 30 |
| Center 3 |  | 34 | 48 | 14 | 96 |
| Center 4 |  |  |  | 58 | 58 |
| Center 5 | 27 | 18 | 39 |  | 84 |
| Center 6 |  | 2 | 100 | 77 | 179 |
| Center 7 | 21 |  | 40 | 45 | 106 |
| Center 8 |  | 3 |  | 8 | 11 |
| **Total cohort** | **93** | **102** | **379** | **313** | **887** |
|  |  |  |  |  |  |
| Rejected after QC | 3 | 5 | 39 | 36 | 83 |
| **WB volumes - approved** | **90** | **97** | **340** | **277** | **804** |
| Approved with remarks | 1 | 1 | 6 | 50 | 58 |
| **GM, WM, CSF, CGM volumes - approved** | **89** | **96** | **334** | **227** | **746** |
|  |  |  |  |  |  |
| No FLAIR available | 35 | 23 | 118 | 82 | 258 |
| Rejected after QC | 15 | 13 | 59 | 83 | 170 |
| **WMH - approved** | **43** | **66** | **202** | **148** | **459** |

Overview of the total REMEMBER cohort (eight centers), including cognitively healthy controls, SCD, MCI, and AD dementia patients (n=887). Volumetric measurements were approved for all volumes (n=746) or MRI scans were completely (n=83) or partly rejected (n=58). Therefore, WB volumes could be analyzed in 804 subjects and the other volumes (GM, WM, CSF, CGM) in 746 subjects. For WMH a FLAIR sequence should be available (n=629) and approved by QC (n=459).

AD, Alzheimer’s disease; CGM, cortical grey matter; CSF, cerebrospinal fluid; GM, grey matter; MCI, mild cognitive impairment; MRI, magnetic resonance imaging; QC, quality control; REMEMBER, retrospective Belgian multi-center MRI biomarker study in dementia; SCD, subjective cognitive decline; WB, whole brain; WM, white matter; WMH, white matter hyperintensities

**Supplementary Table 2.** Detailed information of MRI scanner and image sequences.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **HC (n=93)** | **SCD (n=102)** | **MCI (n=379)** | **AD (n=313)** | **Total (n=887)** |
| **Scanner type** | |  |  |  |  |  |
|  | GE medical systems |  | **21 [95.2]** | **64 [78.1]** | **55 [83.6]** | **140 [82.9]** |
|  | 1.5T |  | 16 [93.8] | 37 [83.8] | 37 [83.8] | 90 [85.6] |
|  | 3.0T |  | 5 [100.0] | 27 [70.4] | 18 [83.3] | 50 [78.0] |
|  | Philips | **27 [96.3]** | **52 [96.2]** | **142 [94.4]** | **87 [71.3]** | **308 [88.3]** |
|  | 1.5T |  | 22 [90.9] | 62 [90.3] | 35 [80.0] | 119 [87.4] |
|  | 3.0T | 27 [96.3] | 30 [100.0] | 80 [97.5] | 52 [65.4] | 189 [88.9] |
|  | SIEMENS | **66 [95.5]** | **29 [89.7]** | **173 [86.7]** | **171 [69.6]** | **439 [81.5]** |
|  | 1.5T |  | 6 [83.3] | 45 [73.3] | 75 [41.3] | 126 [54.8] |
|  | 3.0T | 66 [95.5] | 23 [91.3] | 128 [91.4] | 96 [91.7] | 313 [92.3] |
|  |  |  |  |  |  |  |
| **Field strength** | |  |  |  |  |  |
|  | 1.5T |  | **44 [90.9]** | **144 [83.3]** | **147 [61.2]** | **335 [74.6]** |
|  | 3.0T | **93 [95.7]** | **58 [96.6]** | **235 [91.1]** | **166 [82.5]** | **552 [89.9]** |
|  |  |  |  |  |  |  |
| **Voxel size** |  |  |  |  |  |  |
|  | 3D T1 | 93 [95.7] | 97 [95.9] | 367 [90.7] | 300 [74.3] | **857 [86.1]** |
|  | 2D T1 |  | 5 [60.0] | 12 [8.3] | 13 [30.8] | **30 [26.7]** |
|  | Slice thickness ≤3mm | 93 [95.7] | 95 [97.9] | 333 [95.2] | 266 [78.6] | **787 [90.0]** |
|  | Slice thickness >3mm |  | 7 [42.9] | 46 [40.0] | 47 [38.3] | **100 [38.0]** |
|  |  |  |  |  |  |  |
|  | 3D FLAIR | 45 [95.6] | 23 [95.7] | 85 [90.6] | 58 [93.1] | **211 [92.9]** |
|  | 2D FLAIR | 13 [0.0] | 56 [78.6] | 176 [71.0] | 173 [54.3] | **418 [62.9]** |
|  | Slice thickness ≤3mm |  | 10 [100.0] | 11 [90.9] | 21 [81.0] | **42 [88.1]** |
|  | Slice thickness >3mm | 58 [74.1] | 69 [81.2] | 250 [76.8] | 210 [62.4] | **587 [71.9]** |

Data are numbers and between brackets the percentage (%) of subjects with approved volumetric measurements based on the visual QC. Information about the MRI scanner types, field strengths, and voxel sizes per clinical diagnostic group, including cognitively healthy controls, SCD, MCI and AD dementia patients (n=887). Scanner types were GE medical systems; 1.5T (Signa) or 3.0T (Discovery), Philips; 1.5T (Achieva, Interna) or 3.0T (Achieva, Ingenia), and SIEMENS; 1.5T (Aera, Avanto, Symphony) or 3.0T (Allegra, TrioTim, Skyra, Prisma). The 3T scanners had more approved scans compared to the 1.5T scanners, in which less cognitive impairment showed a higher percentage of approved scans (HC>SCD>MCI>AD dementia). For the different scanner types, this phenomenon was also detected, as less cognitive impairment had a higher percentage of approved scans for both the Philips and SIEMENS scanners. This trend was found for GE scanners as well, however the MCI patients had a lower percentage of approved scans compared to the AD dementia patients. Smaller slice thickness (≤3mm) and 3D T1/FLAIR sequences had a higher percentage of approved scans compared to larger slice thickness (>3mm) and 2D T1/FLAIR sequences. Again, with a higher percentage of approved scans in less cognitive impaired subjects (HC>SCD>MCI>AD dementia).

AD, Alzheimer’s disease; GE, general electric; HC, cognitively healthy controls; MCI, mild cognitive impairment; MRI, magnetic resonance imaging; QC, quality control; REMEMBER, retrospective Belgian multi-center MRI biomarker study in dementia; SCD, subjective cognitive decline; T, Tesla

**Supplementary Table 3.** Overview of analyzes based on T1 or T1 in combination with FLAIR.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **Controls** | **SCD** | **MCI** | **AD** | **Total** |
| **T1 only** |  | **35** | **23** | **108** | **68** | **234** |
|  | **WB volumes - approved** | 34 | 22 | 99 | 61 | 216 |
|  | **GM, WM, CSF, CGM volumes - approved** | 33 | 22 | 96 | 60 | 211 |
|  | **%Rejected scans** | **2.8** | **4.3** | **8.3** | **10.3** | **7.7** |
|  |  |  |  |  |  |  |
| **T1 and FLAIR** | | **58** | **79** | **271** | **245** | **653** |
|  | **WB volumes - approved** | 56 | 75 | 241 | 216 | 588 |
|  | **GM, WM, CSF, CGM volumes - approved** | 56 | 74 | 238 | 167 | 535 |
|  | **%Rejected scans** | **3.4** | **5.1** | **11.1** | **11.8** | **9.9** |
|  |  |  |  |  |  |  |
| **Total cohort** | | **93** | **102** | **379** | **313** | **887** |

Overview of the distribution of image sequences (T1 or T1/FLAIR) per clinical diagnostic group, including cognitively healthy controls, SCD, MCI and AD dementia patients (n=887). Volumetric measurements based on only T1 (n=234) or T1 in combination with FLAIR (n=653). Clinical diagnostic groups were divided bases on the visual QC (see Table 1); approved, approved with remarks, and rejected. The percentage of rejected scans per clinical diagnosis, and also in the total cohort, were for both analyses (T1 only or T1 and FLAIR) comparable. No difference in the usage of T1 only versus T1 in combination with FLAIR was detected, as both analyzing methods had less than 10% rejected scans, respectively 7.7% and 9.9%.

AD, Alzheimer’s disease; CGM, cortical grey matter; CSF, cerebrospinal fluid; GM, grey matter; MCI, mild cognitive impairment; MRI, magnetic resonance imaging; QC, quality control; REMEMBER, retrospective Belgian multi-center MRI biomarker study in dementia; SCD, subjective cognitive decline; WB, whole brain; WM, white matter; WMH, white matter hyperintensities

**Supplementary Table 4.** Cut-off values of the volumetric measurements.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Cut-off (mL)** |  |  |
|  |  | **Versus SCD stage** | **Versus MCI stage** | **Versus AD dementia stage** |
| **WB** |  |  |  |  |
|  | HC | 1390.0 | 1383.2 | 1379.9 |
|  | SCD |  | 1359.9 | 1359.7 |
|  |  |  |  |  |
| **GM** |  |  |  |  |
|  | HC | 816.6 | 809.9 | 814.4 |
|  | SCD |  | 799.5 | 764.3 |
|  |  |  |  |  |
| **WM** |  |  |  |  |
|  | HC | 563.2 | 563.9 | 563.9 |
|  | SCD |  | 605.5 | 550.4 |
|  |  |  |  |  |
| **CSF** |  |  |  |  |
|  | HC | 531.9 | 597.5 | 600.8 |
|  | SCD |  | 617.1 | 637.2 |
|  |  |  |  |  |
| **CGM** |  |  |  |  |
|  | HC | 770.6 | 760.1 | 769.9 |
|  | SCD |  | 733.7 | 723.9 |
|  |  |  |  |  |
| **WMH** |  |  |  |  |
|  | HC | 5.9 | 5.8 | 5.4 |
|  | SCD |  | 8.0 | 5.4 |

Data are cut-off values (mL).

AD, Alzheimer’s disease; CGM, cortical grey matter; CSF, cerebrospinal fluid; GM, grey matter; HC, cognitively healthy controls; MCI, mild cognitive impairment; SCD, subjective cognitive decline; WB, whole brain; WM, white matter; WMH, white matter hyperintensities