Received: 28 June 2024 | Revised: 2 August 2024 | Accepted: 18 August 2024

DOI: 10.1111/jon.13234

JOURNAL OF NEUROIMAGING
THE OFFICIAL JOURNAL OF THE AMERICAN SOCIETY OF NEUROIMAGING
WILEY

CLINICAL INVESTIGATIVE STUDY

Reliability of quantitative magnetic susceptibility imaging metrics for cerebral cortex and major subcortical structures

Maria Agnese Pirozzi[1] | Antonietta Canna[2] | Federica Di Nardo[1] | Mario Sansone[3] | Francesca Trojsi[1] | Mario Cirillo[1] | Fabrizio Esposito[1]

[1]Department of Advanced Medical and Surgical Sciences, University of Campania "Luigi Vanvitelli", Naples, Italy

[2]Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada

[3]Department of Electrical Engineering and Information Technologies, University of Naples "Federico II", Naples, Italy

Correspondence
Fabrizio Esposito, Department of Advanced Medical and Surgical Sciences, University of Campania "Luigi Vanvitelli", Piazza Luigi Miraglia, 2, 80138 Naples, Italy.
Email: fabrizio.esposito@unicampania.it

Funding information
Ministry of University and Research (MUR); NEXTGENERATIONEU (NGEU); National Recovery and Resilience Plan (NRRP); MNESYS, Grant/Award Numbers: PE0000006, DN. 1553 11.10.2022; Ministero dell'Università e della Ricerca, Grant/Award Number: PE0000006 MNESYS

Abstract

Background and purpose: Susceptibility estimates derived from quantitative susceptibility mapping (QSM) images for the cerebral cortex and major subcortical structures are variably reported in brain magnetic resonance imaging (MRI) studies, as average of all (μ_{all}), absolute (μ_{abs}), or positive- (μ_p) and negative-only (μ_n) susceptibility values using a region of interest (ROI) approach.

This pilot study presents a reliability analysis of currently used ROI-QSM metrics and an alternative ROI-based approach to obtain voxel-weighted ROI-QSM metrics (μ_{wp} and μ_{wn}).

Methods: Ten healthy subjects underwent repeated (test-retest) 3-dimensional multi-echo gradient-echo (3DMEGE) 3 Tesla MRI measurements. Complex-valued 3DMEGE images were acquired and reconstructed with slice thicknesses of 1 and 2 mm (3DMEGE1, 3DMEGE2) along with 3DT1-weighted isometric (voxel 1 mm^3) images for independent registration and ROI segmentation.

Agreement, consistency, and reproducibility of ROI-QSM metrics were assessed through Bland-Altman analysis, intraclass correlation coefficient, and interscan and intersubject coefficient of variation (CoV).

Results: All ROI-QSM metrics exhibited good to excellent consistency and test-retest agreement with no proportional bias. Interscan CoV was higher for μ_{all} in comparison to the other metrics where it was below 15%, in both 3DMEGE1 and 3DMEGE2 datasets. Intersubject CoV for μ_{all} and μ_{abs} exceeded 50% in all ROIs.

Conclusions: Among the evaluated ROI-QSM metrics, μ_{all} and μ_{abs} estimates were less reliable, whereas separating positive and negative values (using μ_p, μ_n, μ_{wp}, μ_{wn}) improved the reproducibility within, and the comparability between, subjects, even when reducing the slice thickness. These preliminary findings may offer valuable insights toward standardizing ROI-QSM metrics across different patient cohorts and imaging settings in future clinical MRI studies.

KEYWORDS
MRI, quantitative susceptibility mapping, regional QSM metrics, reliability

INTRODUCTION

Quantitative susceptibility mapping (QSM) is an MRI technique used to map the spatial distribution of bulk magnetic susceptibility, enabling the in-vivo quantification of the content of iron and other chemical species, such as calcium and/or lipids (eg, white matter myelin phospholipids) within cerebral tissues for the study of neurodegeneration processes.[1]

Complex-valued 3D multi-echo gradient-echo (3DMEGE) MRI images capture the effects of the local changes in the induced magnetic field caused by the presence of a magnetic source. The resulting susceptibility variation can be quantified as a positive (ie, the corresponding tissue source is predominantly paramagnetic) or negative (ie, the corresponding tissue source is predominantly diamagnetic) shift in the resonance frequency. Deoxyhemoglobin, hemosiderin, and ferritin are typical paramagnetic sources in the brain, whereas diamagnetic sources may result from calcium deposits and myelinated tissue. Both types of sources contribute to the contrast of QSM images after signal calibration at each voxel. Resolving the variable mixture of paramagnetic and diamagnetic susceptibility sources within the imaging voxel has been previously addressed with special MRI sequences enabling the dissociation of paramagnetic and diamagnetic components at the subvoxel level.[2,3] However, in most QSM-MRI clinical studies, subvoxel modeling and decomposition of QSM images is not possible due to technical limitations, and, therefore, the value at each voxel can only inform about the prevalence of paramagnetic or diamagnetic sources.

Most typically, QSM-based MRI studies report tissue alterations by taking the average of QSM values across all voxels within a region of interest (ROI) which is in turn segmented either from the same QSM image or a separate (registered) 3D-T1-weighted (w) image.[4,5] However, either due to partial volume effects or blooming artifacts in QSM images, segmentation errors could always occur, affecting the regional QSM estimates (hereinafter, referred to as ROI-QSM metrics). Consequently, the simple regional averaging of QSM values across all voxels may penalize the reproducibility of the ROI-QSM metrics, as well as their comparability across different studies or cohorts. Typical solutions to this problem consist of the selective averaging of positive- and negative-only QSM values within the ROI mask,[6–8] although other studies have also reported the average of absolute QSM values.[9,10] Although ROI segmentation methods can always be improved in several alternative ways, for example, by introducing slight modifications to pre-existing procedures[11] or by consolidating the use of QSM-based atlases,[12] and even correcting for partial volume effects would be possible,[13] it remains crucial to explore the reliability of the different approaches to obtain ROI-QSM metrics under typical settings used for generating QSM images in clinical MRI studies.

ROI-based studies were previously conducted to evaluate the reproducibility of subcortical QSM metrics across different MRI scanners,[14] centers,[15] reconstruction methods,[16] field strength, and (sets of) echo times.[17,18] Another study targeted specific cerebral regions under conditions of full-blown neurological pathology.[19] However, a systematic analysis of the test-retest reliability of ROI-QSM metrics commonly used in clinical studies, from QSM images repeatedly acquired in the same cohort, is currently missing. Assessments of ROI-QSM metrics are crucial to standardize QSM-based assessments (also across different patient populations and imaging settings) to obtain more reliable and less variable ROI-QSM metrics, toward possibly more reproducible and comparable conventional QSM-based results. Therefore, this pilot study presents a preliminary test-retest reliability analysis of the currently used ROI-QSM metrics, as well as of an alternative (novel) supra-voxel approach to obtain voxel-weighted ROI-based susceptibility estimates to extract subregions of locally more paramagnetic (ie, positive-only QSM values) and more diamagnetic (ie, negative-only QSM values) by fitting a spatial distribution with multiple Gaussian components. QSM images were obtained from repeated scans (test-retest) in a sample of 10 healthy young adults using the same scanner within a short time (2-4 weeks). Given the importance of QSM in the study of aging and diseased populations, the effect of increasing the slice thickness in the multi-echo sequence (eg, to reduce scan duration) on the reproducibility of the ROI-QSM metrics was also evaluated.

METHODS

Subjects

Ten healthy volunteers (5 females; 31.7±6.63 years; range: 24-46 years) underwent MRI acquisition. An exclusion criterion was the presence of neurological disease to avoid possible pathological bias in the QSM measurements. Written informed consent was signed by each participant before MRI acquisition. The study was conducted according to the Declaration of Helsinki and approved by the Ethics Committee (Protocol nr.591/2018).

MRI acquisitions

MRI data were collected in a test-retest procedure with an average time between the two acquisitions of 33.5±7.46 days (range: 8-55 days) and acquired on a 3 Tesla (3T) MR750 scanner equipped with a 32-channel parallel head coil (General Electric Healthcare, Milwaukee, WI). The imaging protocol included:

- 3D-T1-weighted sequence (gradient-echo sequence inversion recovery prepared fast spoiled gradient recall-echo): repetition time (TR) = 6900 milliseconds, echo time (TE) = 3.0 milliseconds, resolution = 1×1×1 mm^3, matrix size = 256×256, inversion time = 650 milliseconds; duration = 6.49 minutes;
- 3D-multi-echo gradient echo sequence with a slice thickness of 1 mm (3DMEGE1): TR = 47.2 milliseconds, first TE = 3.2 milliseconds, echo spacing = 3.75 milliseconds, number of echoes = 12, resolution = 1×1×1 mm^3, matrix size = 256×256, anterior-posterior

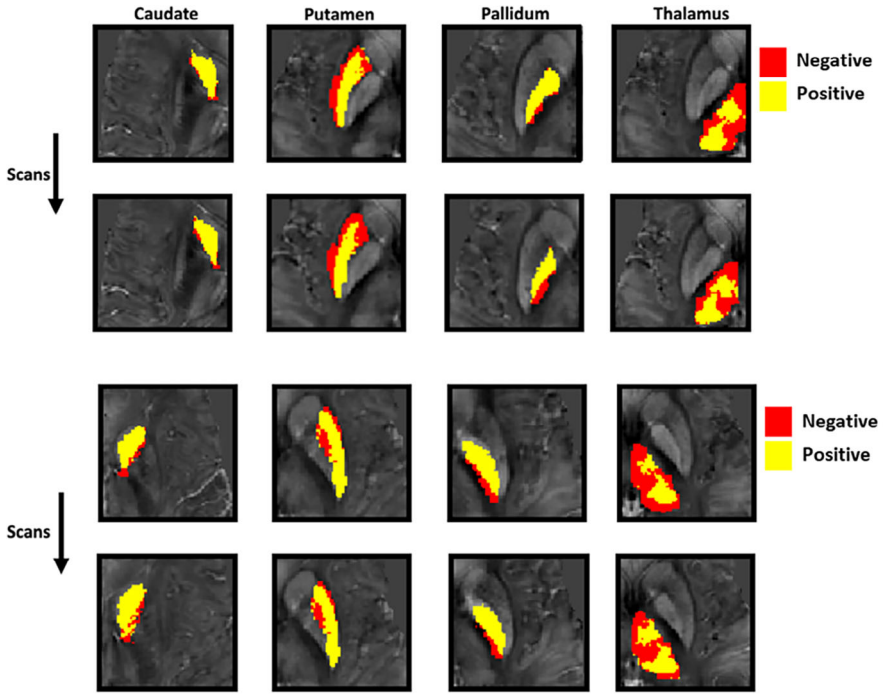

FIGURE 1 Masks of the spatial distribution of negative (red) and positive (yellow) voxels across test and retest (along the arrow from test to retest) in four subcortical regions (upper rows: right hemisphere, bottom rows: left hemisphere) calculated on the QSM map (derived from 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 1 mm) of a representative subject.

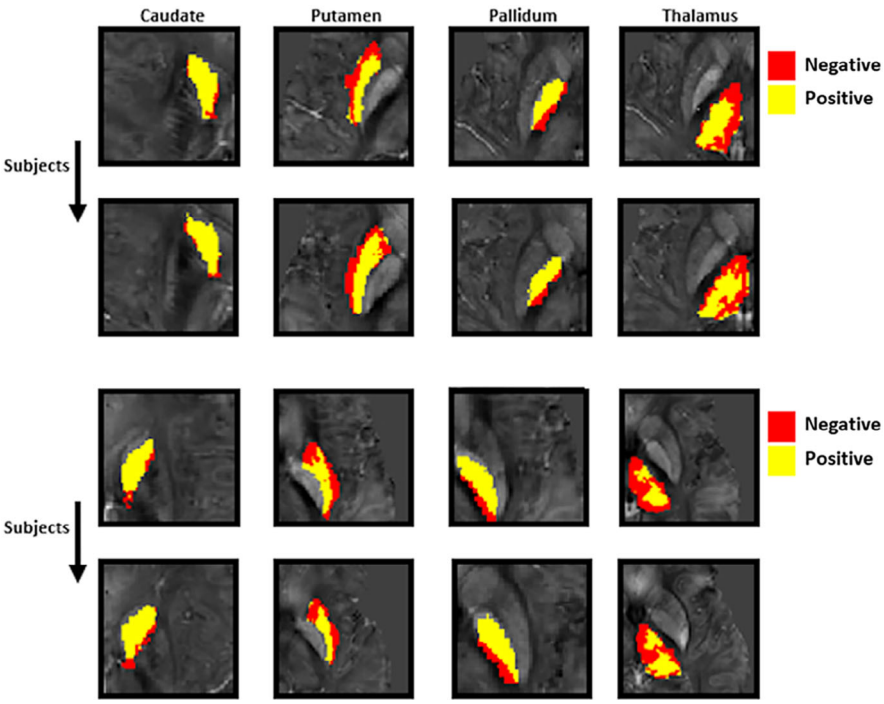

FIGURE 2 Masks of the spatial distribution of negative (red) and positive (yellow) voxels across QSM maps (derived from 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 1 mm, test session) of two representative subjects (along the arrow from subject 1 to subject 2) in four subcortical regions (upper rows: right hemisphere, bottom rows: left hemisphere).

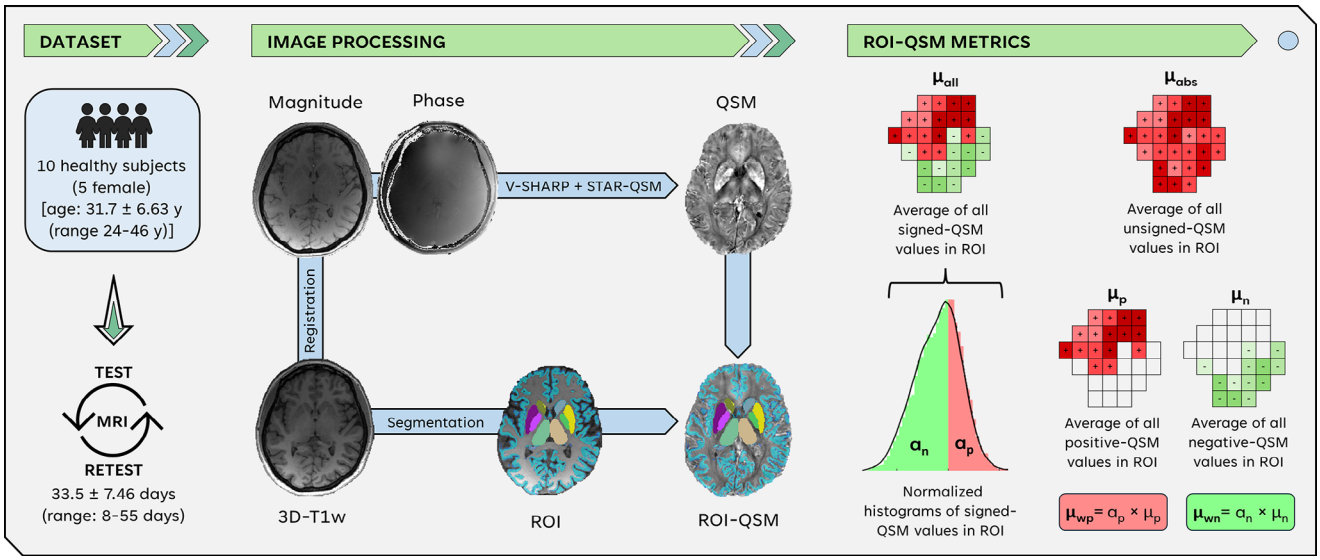

FIGURE 3 Flowchart of the study. The data set consisted of 10 healthy volunteers (5 female; 31.7±6.63; range: 24-46 years), who underwent test and retest (33.5±7.46 days later) MRI acquisitions. Complex-valued 3D-multi echo gradient echo (magnitude and phase) images were computed to obtain QSM maps and 3D-T1weighted isometric (voxel 1 mm^3) images were segmented and registered in the QSM native space. Finally, QSM metrics (μ_{all}, μ_{abs}, μ_p, μ_n, μ_{wp}, μ_{wn}) for each region of interest (ROI), here called ROI-QSM metrics. Abbreviations: 3D-T1w, 3-dimensional-T1-weighted; a_p , weight for the positive distributions; a_n , weight for the negative distributions; STAR-QSM, streaking artifact reduction for QSM; y, years; μ_{all}, average of all signed-QSM values; μ_{abs}, average of absolute-QSM values; μ_p, average of positive-only QSM values; μ_{wp}, weighted-positive average of QSM values; μ_n, average of negative-only QSM values; μ_{wn}, weighted-negative average of QSM values; V-SHARP, sophisticated harmonic artifact reduction for phase data with varying spherical kernel sizes.

phase-encoding direction. Number of signal averages (NEX) = 0.7; duration = 10.18 minutes;

- 3D-multi-echo gradient echo sequence with a slice thickness of 2 mm (3DMEGE2): TR = 47 milliseconds, first TE = 3.0 milliseconds, echo spacing = 3.75 milliseconds, number of echoes = 12, resolution = 1×1×2 mm^3, matrix size = 256×256, anterior-posterior phase-encoding direction. NEX = 0.7; duration = 5.09 minutes.

Image processing

For each complex-valued 3DMEGE acquisition, the magnitude and phase images were computed in MATLAB (v.R2018b, The MathWorks, Inc, https://www.mathworks.com/) to obtain QSM maps via the pipeline implemented in STISuite (v.3.0, Berkeley, CA, https://people.eecs.berkeley.edu/~chunlei.liu/software.html). Magnitude images of the first echo were brain masked through the Brain Extraction Tool (BET2, FMRIB, Oxford, UK, https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET),[20] whereas raw phase images were processed by applying in the order (1) the Laplacian-based method for 3-dimensional phase unwrapping; (2) the sophisticated harmonic artifact reduction for phase data with varying spherical kernel sizes (V-SHARP) method[21] for background field removal; and (3) the streaking artifact reduction for QSM (STAR-QSM) method to obtain effective streaking artifact removal and accurate susceptibility quantification in the computed QSM maps by the dipole convolution.[22] V-SHARP and STAR-QSM combination yielded QSM maps not containing large-scale susceptibility inhomogeneities.[22] QSM values were measured in parts per billion and directly used for comparisons without referencing to any cerebrospinal fluid region, however, setting the reference to the whole-brain susceptibility,[2,22] since it was also previously demonstrated that under this assumption no systematic errors were observed when referencing or not referencing QSM maps to cerebrospinal fluid.[23]

3D-T1w images were processed using FreeSurfer (v.6.0, Martinos Center for Biomedical Imaging, Charlestown, MA, https://freesurfer.net/) using the standard structural image preprocessing and segmentation pipeline via the recon-all procedure. ROI masks of the left and right thalamus, caudate, putamen, and pallidum, and masks of the whole cortex were extracted. Then, the ROI masks of each subject were coregistered to QSM maps in a two-step procedure using FLIRT (FMRIB's Linear Image Registration Tool with 12 degrees of freedom, Oxford, UK, https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT)[24]: (1) the first echo magnitude images were coregistered to the 3D-T1w volumes; (2) the corresponding transformation matrices were inverted and applied to the ROI masks derived from the 3D-T1w volume. The same processing was applied to both the 3DMEGE1 and 3DMEGE2.

ROI-QSM metrics

The overall mean (μ_{all}), the mean of positive-only (μ_p), the mean of negative-only (μ_n), and the mean of absolute QSM values (μ_{abs}) in each ROI were computed in MATLAB. Figures 1 and 2 show the spatial distribution of negative and positive voxels in some subcortical structures in a representative subject.

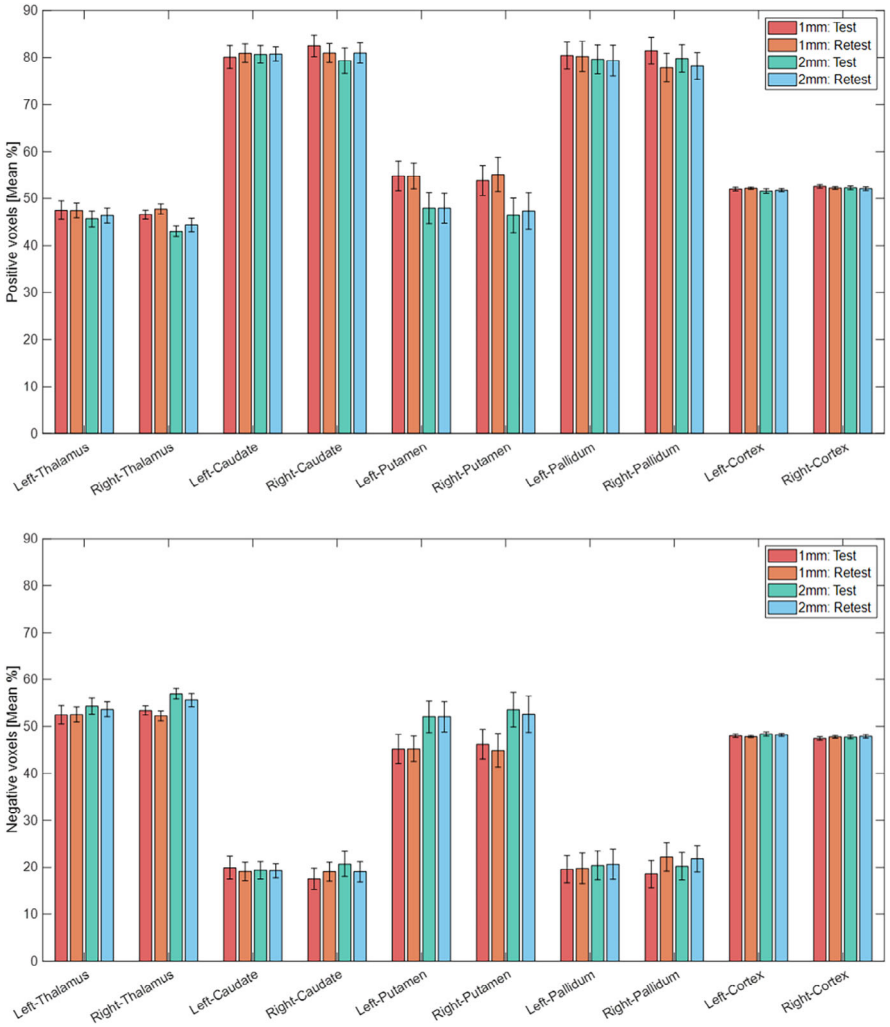

FIGURE 4 Bar graphs representing the average percentage of positive and negative QSM voxels in each region of interest. Bars are shown for both test and retest series acquired using a 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 1 and 2 mm. Abbreviation: mm, millimeters.

In addition, a different estimation approach that in principle avoids the rigid selection criterion of voxel in the ROI and weights the average susceptibility based on the distribution of the voxel values within the ROI was evaluated. The weights for weighted-positive (μ_{wp}) and weighed-negative susceptibility averages (μ_{wn}) were computed by exploiting the normalized histograms of all voxels in the ROI. In this way, the auto-binning procedure ensures uniform bins across the entire range of regional QSM values and a total area under the curve equal to 1. Thus, by summing up the heights of the bar separately for the positive and negative values of the bin centers, two weights for the positive and negative distributions (a_p and a_n) were obtained (whose sum equals to 1). Finally, the corresponding weighted means were calculated as:

$$\mu_{wp} = a_p * \mu_p,$$

$$\mu_{wn} = a_n * \mu_n.$$

The flowchart is summarized in Figure 3.

Statistical analysis

The percentage of positive and negative QSM voxels in each ROI was evaluated to quantify the amount of positive and negative voxels in each ROI and Wilcoxon signed-rank tests were performed to assess within-subject (test-retest, across-ROIs) differences in these percentages for each ROI (for both 3DMEGE1 and 3DMEGE2).

Wilcoxon signed-rank tests were performed to assess the within-subject (test-retest, across-ROIs) and between-subject (test-retest, across-subjects) differences in ROI-QSM metrics derived from both 3DMEGE1 and 3DMEGE2. Then, the overall agreement between measurements over time (test vs. retest) was further evaluated through a Bland-Altman analysis[25] for each ROI-QSM metric.

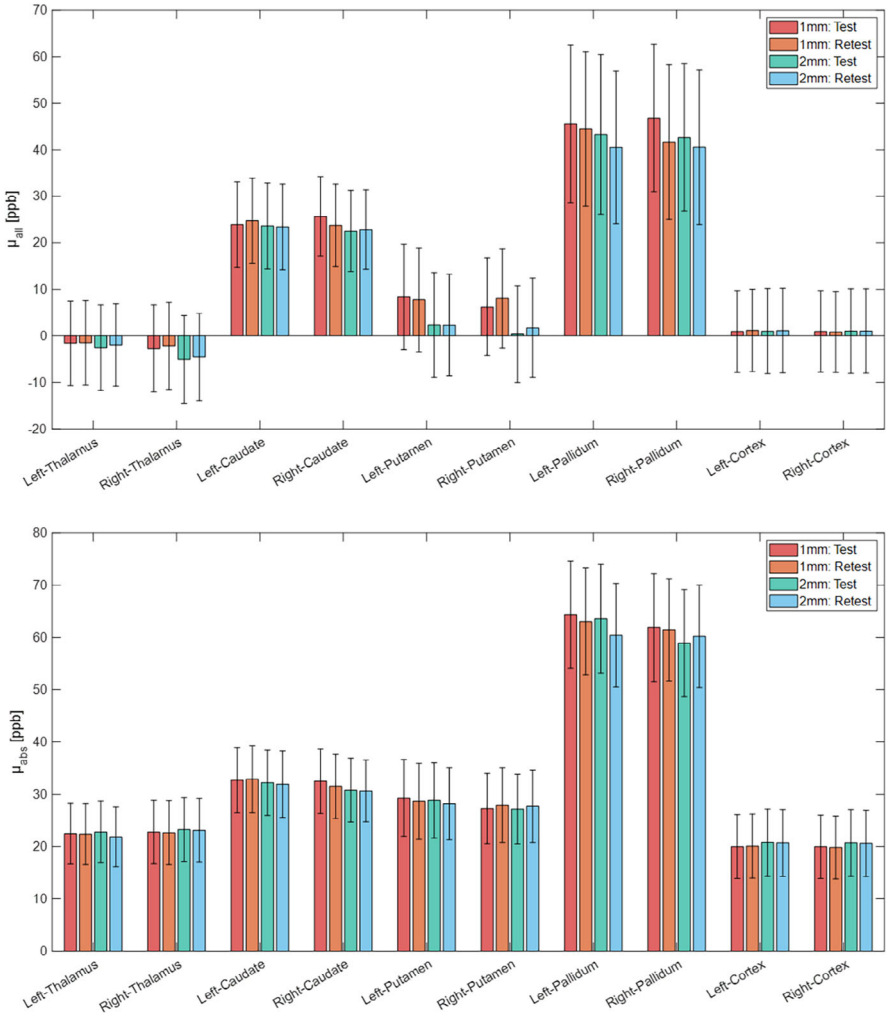

FIGURE 5 Bar graphs representing the average across all subjects of QSM metrics average of all signed-QSM values (μ_{all}) and average of absolute-QSM values (μ_{abs}) for each region of interest. Bars are shown for both test and retest series acquired using a 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 1 and 2 mm. Abbreviations: mm, millimeters; ppb, parts per billion.

For a comprehensive assessment of the ROI-QSM metrics reliability in major subcortical structures, in terms of across-subject consistency of measurements between the two scan sessions (test-retest), and the two resolution configurations (3DMEGE1 vs. 3DMEGE2), the intraclass correlation coefficient (ICC), in the two-way mixed effect model with consistency formulation, was evaluated. ICC provides an index of consistency and can be qualitatively categorized according to the following ranges: poor ICC < 0.40, fair 0.4 ≤ ICC < 0.59, good 0.60 ≤ ICC ≤ 0.75, and excellent consistency ICC > 0.75.[26]

Finally, to assess the variability of the ROI-QSM metrics, the interscan and intersubject coefficients of variations (CoVs) were computed for 3DMEGE1 and 3DMEGE2. The across-subject interscan CoVs were calculated according to the following formula for each ROI-QSM metric:

$$\mathrm{CoV}_{\mathrm{interscan,\ xx}}\,[\%] = 200 * \left(\frac{\mu_{\mathrm{xx,\ 1}} - \mu_{\mathrm{xx,2}}}{\mu_{\mathrm{xx,1}} + \mu_{\mathrm{xx,2}}}\right),$$

where $\mu_{xx,1}$ and $\mu_{xx,2}$ represented the ROI-QSM metric computed from the test and retest QSM maps, respectively.

The intersubject CoVs were computed as:

$$\mathrm{CoV}_{\mathrm{intersubjects,\ xx}}\,[\%] = 100 * \frac{\mathrm{SD}_{\mathrm{xx}}}{\mu_{\mathrm{xx}}},$$

where SD_{xx} and μ_{xx} represented the standard deviation and the mean across subjects computed for ROI-QSM metrics. Nonparametric Wilcoxon tests were also performed to compare intersubject CoVs, as well as to compare interscan CoVs within each series (3DMEGE1 and 3DMEGE2).

ICCs and Bland Altman analyses were computed using the physc and blandr packages in the Rstudio (v.2023.12.0+369, Boston, MA, https://posit.co/products/open-source/rstudio/), whereas CoVs calculation and the nonparametric Wilcoxon signed-rank tests were performed in MATLAB. Statistically significant *p*-value was set at .05 and *p*-values were corrected using false discovery rate.

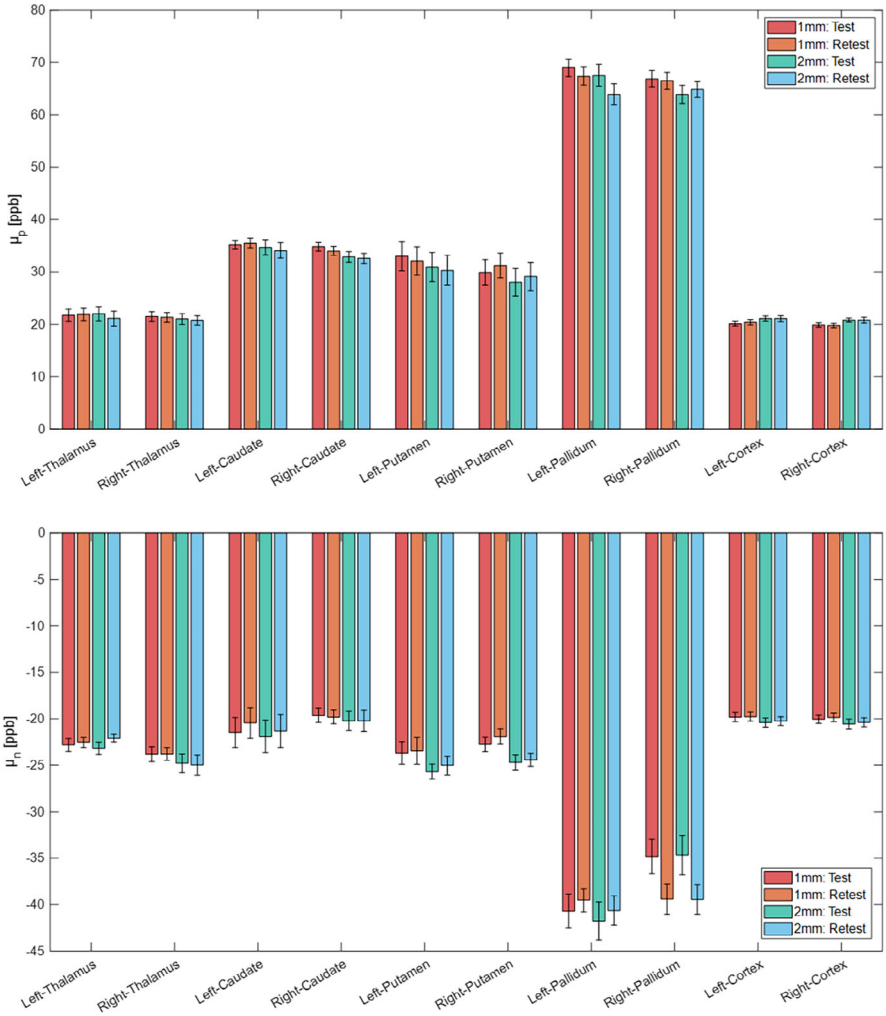

FIGURE 6 Bar graphs representing the average across all subjects of QSM metrics average of positive-only QSM values (μ_p) and average of negative-only QSM values (μ_n) for each region of interest. Bars are shown for both test and retest series acquired using a 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 1 and 2 mm. Abbreviations: mm, millimeters; ppb, parts per billion.

RESULTS

For all subjects, both for 3DMEGE1 and 3DMEGE2, no significant within-subject differences (across ROIs) were found in the percentage of positive and negative QSM voxels from each ROI ($p > .05$ for all ROIs when comparing test vs. retest percentage of positive and negative voxels). Bar graphs showing the mean percentage of positive and negative QSM voxels in ROIs are in Figure 4. Similarly, no significant within-subject differences (across ROIs) of all ROI-QSM metrics ($p > .05$ when comparing test vs. retest metrics per subject), as well as no significant between-subject differences for each ROI ($p > .05$ when comparing test vs. retest metrics per ROI), were revealed. Bar graphs showing the average across all subjects obtained for ROI-QSM metrics are in Figures 5-7.

Bland-Altman analysis between the ROI-QSM metrics (test vs. retest) allowed quantifying the overall agreement, and the regression analysis between the mean and the difference in Bland-Altman plots confirmed the absence of any proportional bias between the compared measures (absolute value of the standardized beta coefficient less than 0.005 and $p > .05$ for all comparisons between ROI-QSM metrics).

Test-retest (over-time, across-subject) consistency of μ_{all} resulted in ICC > 0.75 for all ROIs except for the right thalamus and left cortex when evaluated using 3DMEGE1, whereas on 3DMEGE2, it was > 0.75 for all ROIs. Considering μ_p, μ_{wp} , and μ_{abs} , ICC > 0.75 was found for all ROIs except for the left caudate only when using the 3DMEGE1. For μ_n, ICC > 0.75 was found in all ROIs except for the right caudate, both using 3DMEGE1 and 3DMEGE2, and for the left thalamus using 3DMEGE2, whereas for μ_{wn} in all ROIs except for the right pallidum using 3DMEGE1 and left thalamus using 3DMEGE2. Furthermore, the across-subject reliability as the image resolution varies (3DMEGE1 vs. 3DMEGE2) evaluated for each acquisition session (test or retest) resulted in > 0.75 for all ROIs and regional QSM metrics. In all reported cases where ICC did not exceed the threshold for excellent consistency, it was still > 0.60, which is the threshold for good consistency. ICC was considered also to assess the consistency of ROI-QSM metrics in the same acquisition session (test or retest) when varying acquisition slice

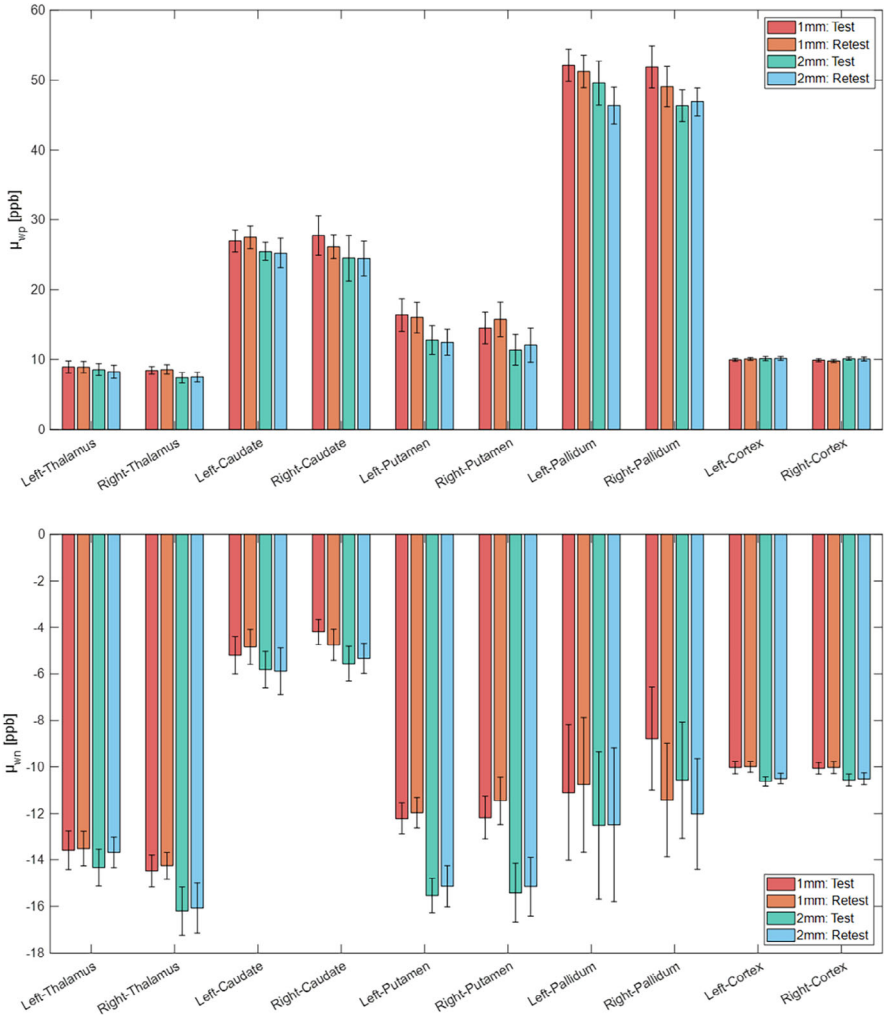

FIGURE 7 Bar graphs representing the average across all subjects of QSM metrics weighted-positive average of QSM values (μ_{wp}) and weighted-negative average of QSM values (μ_{wn}) for each region of interest. Bars are shown for both test and retest series acquired using a 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 1 and 2 mm. Abbreviations: mm, millimeters; ppb, parts per billion.

thickness, resulting in excellent reliability except for μ_{all} in the right cortex (ICC = 0.74) and μ_{wn} in the left thalamus, (ICC = 0.72), however, confirming good consistency also in these cases.

In terms of within-subject reproducibility, the absolute value of the interscan CoVs (%) was evaluated for ROI-QSM metrics and reported in Tables 1 and 2 (for 3DMEGE1 and 3DMEGE2, respectively). No significant differences in the interscan variability between the two series for each ROI-QSM metric were found. Moreover, between-subject reproducibility of all the considered ROI-QSM metrics was assessed in terms of intersubject CoVs (%) and reported in Table 3 and Table 4 (for 3DMEGE1 and 3DMEGE2, respectively). The highest values were obtained for μ_{all}, especially for the thalamus and cortex regions, whereas the lowest values were obtained for μ_{p}, compared to μ_{abs} and μ_{wp}, especially in the caudate and pallidum regions. For 3DMEGE1 and 3DMEGE2, the intersubject variability (across all ROIs) was not significantly different between test-retest sessions (Wilcoxon signed-rank test, $p > .05$) for all ROI-QSM metrics.

DISCUSSION

A pilot study of the reliability of ROI-QSM metrics in the cerebral cortex and major subcortical structures was conducted using a test-retest experimental design over a short temporal interval. 3DMEGE acquisitions were obtained on a sample of 10 healthy volunteers with two different slice thicknesses (1 and 2 mm for 3DMEGE1 and 3DMEGE2, respectively). QSM maps were computed through a widely used calibration approach, and a standard pipeline was applied for the automatic segmentation of cerebral cortex and major subcortical structures on independently acquired 3D-T1w images, after coregistration to QSM maps. This work aimed to evaluate the reliability of the possible derived ROI-QSM metrics. These preliminary assessments should indeed be crucial for choosing ROI-QSM metrics in clinical evaluation and gaining more confidence in QSM-derived measurements in iron or myelin content changes in major (sub)cortical structures, especially when comparing patients with neurological diseases to healthy controls.

TABLE 1 Interscan absolute coefficients of variations for 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 1 mm.

Left (Right) ROI Interscan |CoV| 3DMEGE1	μ_{all}	μ_{abs}	μ_{p}	μ_{wp}	μ_{n}	μ_{wn}
Thalamus	7.71% (22.49%)	0.36% (0.52%)	0.71% (0.83%)	1.13% (0.08%)	0.29% (1.52%)	0.50% (1.54%)
Caudate	3.37% (7.73%)	0.57% (3.18%)	0.81% (2.30%)	4.89% (0.84%)	1.89% (6.03%)	7.09% (12.47%)
Putamen	7.82% (25.90%)	2.39% (2.21%)	2.85% (4.33%)	1.13% (3.79%)	2.18% (7.84%)	2.09% (6.02%)
Pallidum	2.33% (11.80%)	2.07% (0.74%)	2.38% (0.58%)	2.88% (12.37%)	1.66% (5.57%)	3.10% (26.13%)
Cortex	21.23% (10.65%)	0.52% (0.76%)	1.25% (0.57%)	0.24% (0.99%)	1.26% (1.27%)	0.33% (0.30%)

Abbreviations: 3DMEGE1, 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 1 mm; CoV, coefficient of variations; ROI, region of interest; μ_{all}, average of all signed-QSM values; μ_{abs}, average of absolute-QSM values; μ_{p}, average of positive-only QSM values; μ_{wp}, weighted-positive average of QSM values; μ_{n}, average of negative-only QSM values; μ_{wn}, weighted-negative average of QSM values.

TABLE 2 Interscan absolute coefficients of variations for 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 2 mm.

Left (Right) ROI Interscan |CoV| 3DMEGE2	μ_{all}	μ_{abs}	μ_{p}	μ_{wp}	μ_{n}	μ_{wn}
Thalamus	25.67% (10.62%)	4.23% (0.55%)	4.12% (1.34%)	4.87% (0.78%)	3.58% (1.51%)	4.70% (0.77%)
Caudate	0.88% (1.39%)	0.95% (0.53%)	1.63% (0.86%)	2.62% (0.04%)	0.86% (0.20%)	1.17% (4.19%)
Putamen	1.27% (128.90%)	2.24% (1.87%)	1.96% (3.80%)	2.55% (1.15%)	2.51% (5.95%)	2.55% (1.75%)
Pallidum	6.67% (5.11%)	5.13% (2.19%)	5.52% (1.56%)	2.84% (12.97%)	6.69% (1.26%)	0.24% (12.90%)
Cortex	12.65% (0.91%)	0.43% (0.47%)	0.04% (0.07%)	0.86% (0.51%)	0.26% (0.48%)	1.14% (0.51%)

Abbreviations: 3DMEGE2, 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 2 mm; CoV, coefficient of variations; ROI, region of interest; μ_{all}, average of all signed-QSM values; μ_{abs}, average of absolute-QSM values; μ_{p}, average of positive-only QSM values; μ_{wp}, weighted-positive average of QSM values; μ_{n}, average of negative-only QSM values; μ_{wn}, weighted-negative average of QSM values.

TABLE 3 Intersubject absolute coefficients of variations for 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 1 mm.

Left (Right) ROI Intersubject |CoV| 3DMEGE1	μ_{all}		μ_{abs}		μ_{p}		μ_{wp}		μ_{n}		μ_{wn}	
	Test	Retest	Test	Retest	Test	Retest	Test	Retest	Test	Retest	Test	Retest
Thalamus	1798.20% (1096.65%)	1946.65% (1373.06%)	81.65% (84.49%)	82.42% (85.23%)	17.64% (14.23%)	17.76% (13.94%)	9.85% (10.42%)	8.12% (9.03%)	30.11% (20.63%)	28.47% (24.77%)	19.33% (15.03%)	17.54% (12.89%)
Caudate	121.20% (105.75%)	116.89% (117.57%)	60.00% (59.80%)	61.16% (61.84%)	7.11% (7.29%)	8.02% (7.92%)	23.92% (12.26%)	25.23% (11.99%)	18.89% (32.16%)	18.62% (20.80%)	48.87% (40.79%)	48.77% (43.93%)
Putamen	427.01% (533.20%)	456.11% (419.39%)	79.44% (78.51%)	80.49% (81.10%)	26.26% (25.69%)	26.64% (24.05%)	15.99% (10.91%)	19.81% (11.90%)	44.63% (49.61%)	42.95% (49.99%)	17.16% (23.96%)	17.40% (28.38%)
Pallidum	117.60% (106.77%)	117.97% (126.24%)	50.45% (52.72%)	51.19% (50.16%)	7.62% (7.52%)	8.16% (7.75%)	14.20% (16.59%)	9.99% (13.09%)	13.99% (18.28%)	14.42% (18.73%)	83.06% (79.58%)	85.05% (67.27%)
Cortex	2898.90% (2886.26%)	2354.92% (3186.79%)	97.17% (96.60%)	97.18% (96.59%)	6.99% (6.94%)	7.84% (6.80%)	8.04% (7.14%)	7.86% (7.35%)	6.56% (6.53%)	6.05% (6.66%)	8.41% (7.76%)	7.48% (8.18%)

Abbreviations: 3DMEGE1, 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 1 mm; CoV, coefficient of variations; ROI, region of interest; μ_{all}, average of all signed-QSM values; μ_{abs}, average of absolute-QSM values; μ_{p}, average of positive-only QSM values; μ_{wp}, weighted-positive average of QSM values; μ_{n}, average of negative-only QSM values; μ_{wn}, weighted-negative average of QSM values.

No significant differences in the percentage of positive and negative voxels were found between test and retest acquisitions, for both 3DMEGE1 and 3DMEGE2, in any of the ROIs delineated by automatic segmentation. Furthermore, no significant differences were detected between ROI-QSM metric averages across repeated measurements, both within-subject (all ROIs) and between-subject (for each ROI). The Bland-Altman analysis demonstrated the absence of proportional biases (between repeated measurements) in all comparisons and of systematic errors in the procedure used to extract the ROI-QSM metrics in all data sets. Particularly, this suggests how the use of a fully automated procedure for the extraction of ROI-QSM metrics avoids the subjective biases usually implied by manual ROI definition.

JOURNAL OF NEUROIMAGING
THE OFFICIAL JOURNAL OF THE AMERICAN SOCIETY OF NEUROIMAGING

TABLE 4 Intersubject absolute coefficients of variations for 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 2 mm.

Left (Right) ROI Intersubject |CoV| 3DMEGE2	μ_{all}		μ_{abs}		μ_{p}		μ_{wp}		μ_{n}		μ_{wn}	
	Test	Retest	Test	Retest	Test	Retest	Test	Retest	Test	Retest	Test	Retest
Thalamus	1151.72% (586.31%)	1437.14% (649.69%)	81.77% (83.98%)	83.10% (83.76%)	19.60% (15.51%)	21.39% (13.54%)	8.79% (12.80%)	6.24% (13.66%)	31.43% (32.35%)	35.11% (32.35%)	17.41% (20.18%)	15.44% (21.11%)
Caudate	123.19% (122.18%)	123.89% (118.14%)	61.28% (62.92%)	63.23% (61.21%)	12.90% (9.95%)	13.69% (9.53%)	25.45% (16.29%)	26.67% (17.96%)	16.20% (42.20%)	26.19% (32.44%)	42.80% (41.98%)	53.86% (38.36%)
Putamen	1518.78% (1786.86%)	1492.68% (1928.24%)	79.10% (77.99%)	77.29% (79.31%)	28.75% (30.40%)	29.90% (29.02%)	9.79% (10.52%)	12.81% (12.71%)	51.61% (60.29%)	47.79% (64.55%)	15.09% (26.05%)	18.48% (26.36%)
Pallidum	125.24% (117.23%)	128.16% (129.56%)	51.93% (54.89%)	51.90% (51.64%)	9.84% (8.70%)	10.00% (7.19%)	15.47% (19.33%)	12.28% (12.71%)	20.20% (15.65%)	18.06% (13.42%)	80.20% (74.84%)	83.70% (62.70%)
Cortex	2874.38% (2745.79%)	2521.09% (2710.49%)	97.61% (97.01%)	97.57% (97.17%)	7.66% (5.50%)	8.38% (8.02%)	7.99% (8.27%)	7.67% (7.87%)	8.95% (6.83%)	8.61% (8.73%)	6.19% (7.79%)	6.49% (7.76%)

Abbreviations: 3DMEGE2, 3-dimensional-multi-echo gradient echo sequence with a slice thickness of 2 mm; CoV, coefficient of variations; ROI, region of interest; μ_{all}, average of all signed-QSM values; μ_{abs}, average of absolute-QSM values; μ_{p}, average of positive-only QSM values; μ_{wp}, weighted-positive average of QSM values; μ_{n}, average of negative-only QSM values; μ_{wn}, weighted-negative average of QSM values.

Nonetheless, the manual definition can be often more accurate in outlining an ROI on a given individual subject anatomy but is certainly more time-consuming and bias-prone.[27,28]

Test-retest across-subject consistency of ROI-QSM metrics, as evaluated through ICC, turned out to be excellent (> 0.75) for most ROIs and otherwise good (> 0.6), for both data sets (3DMEGE1 and 3DMEGE2), that is, regardless of the slice thickness. Test-retest variability, as evaluated through interscan and intersubject CoV, was higher for μ_{all}, in comparison to the other ROI-QSM metrics, for both 3DMEGE1 and 3DMEGE2 data sets. Thus, this metric is definitely less reliable, especially in ROIs expectedly more heterogeneous in terms of susceptibility, such as thalamus, putamen, and cortex. Particularly, the higher intersubject variability for μ_{all} (but also μ_{abs}) in thalamus should be ascribed to the larger variability in the distribution of magnetic susceptibility across the many (smaller) thalamic nuclei causing a focal prevalence of the relatively myelin-rich contiguous gray matter (GM) and white matter (WM) subregions within the internal and external intermedullary lamina and the consequent combination of different paramagnetic and diamagnetic sources contributing to both positive- and negative-valued QSM estimates.[29] Indeed, μ_{abs} and μ_{p} metrics have been previously proposed as better (positive definite) alternatives to μ_{all} due to the sign concordance with the (positive definite) R2* parameter and the reduced impact of blooming artifacts.[8,9] However, the intersubject variability of μ_{abs} was still above 50% in all ROIs, suggesting that this metric may not be as reproducible a metric as hoped. In contrast, here we show that μ_{p} had the lowest variability (ie, less than 20% in all ROIs except putamen), and, therefore, it should be considered as a more reliable positive definite metric for regional analysis only of the prevalence of paramagnetic sources (including iron) in the bulk susceptibility of QSM voxels.

Note instead that, compared to μ_{p}, the newly introduced μ_{wp} metric showed an even reduced variability for the putamen and thalamus ROIs, albeit also an increased variability in the caudate and pallidum. Similar (complementary) evidence was observed in the same ROIs when μ_{n} and μ_{wn} were compared, suggesting that the sign weighting of the susceptibility values could be a reasonable (or preferable) choice when extracting ROI-QSM metrics for putamen and thalamus.

For similar reasons, the cerebral cortex intersubject CoV was higher for μ_{all} and μ_{abs} metrics. Indeed, compared to the subcortical ROIs, the cortex is a much larger region, in which despite the prevalence of GM voxels, the contribution of negative voxels is certainly not negligible, especially at the interface between GM and WM.[30] Thus, the sole aspect of reducing the blooming effects[9] may not justify the use of this metric for clinical studies. On the other hand, intersubject CoV for μ_{p}, μ_{wp}, μ_{n}, and μ_{wn} were lower, showing how the separation of positive- and negative-valued QSM estimates (μ_{p}, μ_{n}), eventually with distribution-based weighting (μ_{wp}, μ_{wn}) may lead to more reliable ROI-QSM metrics and possibly better surrogate estimates of the prevalence of paramagnetic sources (including iron) or diamagnetic sources (including myelin) content in ROI voxels when appropriate modeling for source separation[2,3] is not applicable. Moreover, positive and negative QSM estimates should be intrinsically less affected by shadow artifacts induced by the anisotropy of the WM voxels adjacent to cortical GM voxels, which may result in unrealistically high negative values within the GM and deep GM.[31] Nonetheless, considering that the cerebral cortex is composed of myelinated and unmyelinated axons mixed with cell bodies of billions of neurons, whose dendrites would mostly be arranged within approximately parallel thin layers, a far more accurate characterization of the susceptibility of these layers would certainly benefit from submillimeter native image resolution of the scans which is currently only compatible with ultra-high field (7T or higher) MRI.

The left/right differences in interscan CoVs (particularly observable for the μ_{all} metric) could be due to non-side-specific artifacts,[32,33] such as residual background field variations, residual nonlocal distortions induced by susceptibility, or even small involuntary movements of the subject.[34] Residual variations may affect the two hemispheres differently during test and retest acquisitions, while small-scale motion may

introduce spurious phase fluctuations, as well as small image shifts, that could generate inconsistencies between the measured field and the magnetostatic of the susceptibility inversion procedure.[34] Such artifacts may influence measurements unevenly between the right and left sides, resulting in the left/right CoVs differences. Furthermore, although accurate and reproducible automatic segmentation was assessed, the delineation of ROI boundaries between the two sides may not be perfectly replicable for both sides between test and retest acquisitions,[35] contributing to the observed left/right variations in interscan CoVs.

All the above considerations on the reliability (ICC) and variability (CoV) of ROI-QSM metrics held true for both 3DMEGE1 and 3DMEGE2 data sets. This means it is possible to increase the resolution of the QSM images (eg, by reducing the acquisition slice thickness), thereby reducing the partial volume effects, without further impairing the reliability of the ROI-QSM metrics, despite the reduction in the signal-to-noise. Alternatively, it is possible to reduce the resolution of the reconstructed QSM images (eg, by increasing the slice thickness), thereby reducing the scan time, without further impairing the reliability of the extracted metrics, despite the increased partial volume effects. In the future, a combination of the proposed (data-driven) supra-voxel decomposition of QSM components with a suitable (model-driven) subvoxel decomposition of the two types of physical QSM sources will possibly allow clarifying the actual impact of slice thickness on partial volume effects. Although this could not be evaluated in more detail in the present study, it should be also considered that slice thickness could be differentially relevant for some of the studied ROIs and that MR imaging resources are usually limited and expensive, thereby long scan times can be inconvenient for the patient, sometimes causing more motion artifacts in the images.[36]

This pilot study has limitations. First, it was focused on the reliability of QSM metrics commonly used in clinical 3T-MRI studies for indexing susceptibility in major subcortical ROIs and in the cerebral cortex. Thus, for some of the larger regions, such as the thalamus and cortex, it may need to be further expanded by considering different parcellations. Moreover, the variability of the structures at the interface with the skull base should be further investigated. This area usually generates artifacts and signal loss, especially in the lower regions of the brain, including, for example, the temporal and inferior frontal lobes, as well as in the entorhinal cortex and the hippocampus. However, the choice of the optimal parcellation and the strategy for eventually combining multiple ROI-QSM estimates across several parcels into one ROI-QSM metric, represents a more general issue for quantitative MRI analyses which goes beyond the scope of this work. Second, this study only analyzed 10 subjects with only one re-test repetition. Therefore, the statistical power was relatively small and the results will need to be further confirmed on larger sample sizes (and with more re-test sessions), eventually extending the same analysis on different cohorts of healthy participants (eg, across different age ranges), as well as to selected groups of neurological patients. Nonetheless, this reliability analysis of ROI-QSM metrics is still robust given the test-retest and double slice-thickness measurements which effectively quadrupled the data sets available for the statistical analyses.

Despite the limitations discussed, these preliminary assessments may offer valuable suggestions toward standardizing QSM-based evaluations across different (and larger) patient populations and imaging settings to obtain more reliable and less variable ROI-QSM metrics (μ_p and μ_n, or also the weighted ones), in the hope to make future results based on conventional QSM, more reproducible and comparable.

ACKNOWLEDGMENTS

Work supported by #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006)—A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022).

CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflict of interest.

ORCID

Maria Agnese Pirozzi https://orcid.org/0000-0002-4632-3513
Fabrizio Esposito https://orcid.org/0000-0002-5099-9786

REFERENCES

1. Fushimi Y, Nakajima S, Sakata A, et al. Value of quantitative susceptibility mapping in clinical neuroradiology. J Magn Reson Imaging. 2024;59:1914-29.
2. Chen J, Gong N-J, Chaim KT, et al. Decompose quantitative susceptibility mapping (QSM) to sub-voxel diamagnetic and paramagnetic components based on gradient-echo MRI data. Neuroimage. 2021;242:118477.
3. Shin H-G, Lee J, Yun YH, et al. χ-Separation: magnetic susceptibility source separation toward iron and myelin mapping in the brain. Neuroimage. 2021;240:118371.
4. Kiersnowski OC, Winston GP, Caciagli L, et al. Quantitative susceptibility mapping identifies hippocampal and other subcortical grey matter tissue composition changes in temporal lobe epilepsy. Hum Brain Mapp. 2023;44:5047-64.
5. Pontillo G, Petracca M, Monti S, et al. Unraveling deep gray matter atrophy and iron and myelin changes in multiple sclerosis. AJNR Am J Neuroradiol. 2021;42:1223-30.
6. Coffman CH, White R, Subramanian K, et al. Quantitative susceptibility mapping of both ring and non-ring white matter lesions in relapsing remitting multiple sclerosis. Magn Reson Imaging. 2022;91:45-51.
7. Zachariou V, Bauer CE, Pappas C, et al. High cortical iron is associated with the disruption of white matter tracts supporting cognitive function in healthy older adults. Cerebral Cortex. 2023;33:4815-28.
8. Zachariou V, Bauer CE, Powell DK, et al. Ironsmith: an automated pipeline for QSM-based data analyses. Neuroimage. 2022;249:118835.
9. Betts MJ, Acosta-Cabronero J, Cardenas-Blanco A, et al. High-resolution characterisation of the aging brain using simultaneous quantitative susceptibility mapping (QSM) and R2* measurements at 7T. Neuroimage. 2016;138:43-63.
10. Thomas GEC, Leyland LA, Schrag A-E, et al. Brain iron deposition is linked with cognitive severity in Parkinson's disease. J Neurol Neurosurg Psychiatry. 2020;91:418-25.
11. Pirozzi MA, Tranfa M, Tortora M, et al. A polynomial regression-based approach to estimate relaxation rate maps suitable for multiparametric segmentation of clinical brain MRI studies in multiple sclerosis. Comput Methods Programs Biomed. 2022;223:106957.

12. Zhang Y, Wei H, Cronin MJ, et al. Longitudinal atlas for normative human brain development and aging over the lifespan using quantitative susceptibility mapping. Neuroimage. 2018;171:176-89.
13. McDaniel P, Bilgic B, Fan A, et al. Mitigation of partial-volume effects in susceptibility-based oxygenation measurements by joint utilization of magnitude and phase (JUMP). Magn Reson Med. 2017;77:1713-27.
14. Liu M, Zhao S, Chen Z. Interscanner reproducibility of volumetric quantitative susceptibility mapping about cerebral subcortical gray nuclei at different MR vendors with the same magnetic strength. Brain Behav. 2024;14:e3473.
15. Lancione M, Bosco P, Costagli M, et al. Multi-centre and multi-vendor reproducibility of a standardized protocol for quantitative susceptibility mapping of the human brain at 3T. Phys Med. 2022;103:37-45.
16. Santin MD, Didier M, Valabrègue R, et al. Reproducibility of R2 * and quantitative susceptibility mapping (QSM) reconstruction methods in the basal ganglia of healthy subjects. NMR Biomed. 2017;30:e3491.
17. Lancione M, Donatelli G, Cecchi P, et al. Echo-time dependency of quantitative susceptibility mapping reproducibility at different magnetic field strengths. Neuroimage. 2019;197:557-64.
18. Bordin V, Pirastru A, Bergsland N, et al. Optimal echo times for quantitative susceptibility mapping: a test-retest study on basal ganglia and subcortical brain nuclei. Neuroimage. 2023;278:120272.
19. Reeves JA, Mohebbi M, Zivadinov R, et al. Reliability of paramagnetic rim lesion classification on quantitative susceptibility mapping (QSM) in people with multiple sclerosis: single-site experience and systematic review. Mult Scler Relat Disord. 2023;79:104968.
20. Smith SM. Fast robust automated brain extraction. Hum Brain Mapp. 2002;17:143-55.
21. Li W, Avram AV, Wu B, et al. Integrated Laplacian-based phase unwrapping and background phase removal for quantitative susceptibility mapping. NMR Biomed. 2014;27:219-27.
22. Wei H, Dibb R, Zhou Y, et al. Streaking artifact reduction for quantitative susceptibility mapping of sources with large dynamic range. NMR Biomed. 2015;28:1294-303.
23. Li W, Wu B, Batrachenko A, et al. Differential developmental trajectories of magnetic susceptibility in human brain gray and white matter over the lifespan. Hum Brain Mapp. 2014;35:2698-713.
24. Jenkinson M, Bannister P, Brady M, et al. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage. 2002;17:825-41.
25. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8:135-60.
26. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. Am J Ment Defic. 1981;86:12-137.
27. Chu R, Hurwitz S, Tauhid S, et al. Automated segmentation of cerebral deep gray matter from MRI scans: effect of field strength on sensitivity and reliability. BMC Neurol. 2017;17:172.
28. Fan W, Wang X, Zhang X, et al. Investigating optimal echo times for quantitative susceptibility mapping of basal ganglia nuclei in the healthy brain. Curr Med Imaging. 2020;16:991-96.
29. Madden DJ, Merenstein JL. Quantitative susceptibility mapping of brain iron in healthy aging and cognition. Neuroimage. 2023;282:120401.
30. Wu B, Li W, Guidon A, et al. Whole brain susceptibility mapping using compressed sensing. Magn Reson Med. 2012;67:137-47.
31. Kee Y, Liu Z, Zhou L, et al. Quantitative susceptibility mapping (QSM) algorithms: mathematical rationale and computational implementations. IEEE Trans Biomed Eng. 2017;64:2531-45.
32. Fang J, Bao L, Li X, et al. Background field removal for susceptibility mapping of human brain with large susceptibility variations. Magn Reson Med. 2019;81:2025-37.
33. Deistung A, Schweser F, Reichenbach JR. Overview of quantitative susceptibility mapping. NMR Biomed. 2017;30:e3569.
34. Mattern H, Sciarra A, Lüsebrink F, et al. Prospective motion correction improves high-resolution quantitative susceptibility mapping at 7T. Magn Reson Med. 2019;81:1605-19.
35. Kecskemeti SR, Alexander AL. Test-retest of automated segmentation with different motion correction strategies: a comparison of prospective versus retrospective methods. Neuroimage. 2020;209:116494.
36. Plenge E, Poot DHJ, Bernsen M, et al. Super-resolution methods in MRI: can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time? Magn Reson Med. 2012;68:1983-93.

How to cite this article: Pirozzi MA, Canna A, Nardo FD, et al. Reliability of quantitative magnetic susceptibility imaging metrics for cerebral cortex and major subcortical structures. J Neuroimaging. 2024;34:720–31.
https://doi.org/10.1111/jon.13234