Test–Retest Reliability of the Brain Metabolites GABA and Glx With JPRESS, PRESS, and MEGA-PRESS MRS Sequences in vivo at 3T

Arwa Baeshen, PhD,[1,2] Patrik O. Wyss, PhD,[3,4,5] Anke Henning, PhD,[3,4] Ruth L. O'Gorman, PhD,[6] Marco Piccirelli, PhD,[1] Spyridon Kollias, MD,[1] and Lars Michels, PhD[1*]

Background: The optimization of magnetic resonance spectroscopy (MRS) sequences allows improved diagnosis and prognosis of neurological and psychological disorders. Thus, to assess the test–retest and intersequence reliability of such MRS sequences in quantifying metabolite concentrations is of clinical relevance.
Purpose: To evaluate the test–retest and intersequence reliability of three MRS sequences to estimate GABA and Glx = Glutamine+Glutamate concentrations in the human brain.
Study Type: Prospective.
Subjects: Eighteen healthy participants were scanned twice (range: 1 day to 1 week between the two sessions) with identical protocols.
Field Strength/Sequence: 3T using a 32-channel SENSE head coil in the PCC region; PRESS, JPRESS, and MEGA-PRESS sequences.
Assessment: Metabolite concentrations were estimated using LCModel (for PRESS and MEGA-PRESS) and ProFit2 (for JPRESS).
Statistical Tests: The test–retest reliability was evaluated by Wilcoxon signed-rank tests, Pearson's *r* correlation coefficients, intraclass-correlation coefficients (ICC), coefficients of variation (CV), and by Bland–Altman (BA) plots. The intersequence reliability was assessed with Wilcoxon signed-rank tests, Pearson's *r* correlation coefficients, and BA plots.
Results: For GABA, only the MEGA-PRESS sequence showed a moderate test–retest correlation ($r = 0.54$, ICC = 0.5, CV = 8.8%) and the BA plots indicated good agreement ($P > 0.05$) for all sequences. JPRESS provided less precise results and PRESS was insensitive to GABA. For Glx, the *r* and ICC values for PRESS ($r = 0.87$, ICC = 0.9, CV = 2.9%) and MEGA-PRESS ($r = 0.70$, ICC = 0.7, CV = 5.3%) reflect higher correlations, compared with JPRESS ($r = 0.39$, ICC = 0.4, CV = 20.1%).
Data Conclusion: MEGA-PRESS and JPRESS are suitable for the reliable detection of GABA, the first being more precise. The three sequences included in the study can measure Glx concentrations.
Level of Evidence: 2
Technical Efficacy: Stage 1

J. MAGN. RESON. IMAGING 2019.

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.26921

Received May 14, 2019, Accepted for publication Aug 15, 2019.

*Address reprint requests to: L.M., Department of Neuroradiology, University Hospital Zurich, Frauenklinikstrasse 10, 8091 Zurich, Switzerland. E-mail: lars.michels@usz.ch
The first two authors contributed equally to this work.

From the [1]Department of Neuroradiology, University Hospital Zurich, Zurich, Switzerland; [2]Department of Radiological Sciences, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia; [3]Institute for Biomedical Engineering, University and ETH Zurich, Zurich, Switzerland; [4]Max Planck Institute for Biological Cybernetics, Tubingen, Germany; [5]Department of Radiology, Swiss Paraplegic Centre, Nottwil, Switzerland; and [6]MR Centre, Children's University Hospital, Zurich, Switzerland

Additional supporting information may be found in the online version of this article

MAGNETIC RESONANCE SPECTROSCOPY (MRS) allows noninvasive quantification of the concentrations of various brain metabolites in vivo, including γ-aminobutyric acid (GABA), glutamine (Gln), glutamate (Glu), the sum of Glu and Gln (Glx), N-acetyl aspartate (NAA), N-acetyl aspartyl glutamate (NAAG), glutathione (GSH), creatine (Cr), myo-inositol (mI), and choline-containing compounds (Cho) at 3T.[1–3]

The link between altered metabolite concentrations and pathologies is the central topic of an extensive number of clinical investigations. Variations in metabolite concentrations provide information complementary to anatomical magnetic resonance imaging (MRI) and are used extensively as biomarkers to detect, monitor, and predict the outcome of various central nervous system diseases (eg, neurological, oncological, and psychiatric).[4,5] MRS provides, noninvasively and in vivo, metabolic information at the cellular level, thus allowing for the monitoring of disease-related alterations before changes become apparent on morphological images.[5] A precise and reproducible choice of the region of interest (ROI; ie, voxel) is important, since the concentrations of most metabolites vary significantly in different areas of the brain. This is epitomized in the case of GABA, whose concentration is higher in the gray matter (GM) as compared with the white matter (WM), while being negligible in the cerebrospinal fluid (CSF).[2] From a methodological perspective, spectral overlap in MRS is a notable limiting factor, since it hinders the distinction of different metabolites located in the same frequency range. Moreover, other challenging aspects include low signal intensity, due to low concentration of certain metabolites, and *J*-coupling and chemical shift displacement effects.[6,7] For example, GABA has a low concentration and its signal overlaps with signals from, eg, Gln and NAA.

To overcome these challenges and distinguish the metabolites reliably within a ROI, besides the conventional and widely available PRESS (Point RESolved Spectroscopy), customized sequences have been developed, such as JPRESS (*J*-resolved spectroscopy) and MEGA-PRESS (MEscher-GArwood Point RESolved Spectroscopy). The choice of the sequence depends on the type of investigation, the available scanning time, and the local resources in terms of expertise, hardware, and software. MEGA-PRESS is a sequence that uses *J*-editing pulses and is designed specifically for the detection of low-concentration *J*-coupled metabolites. Conversely, JPRESS allows for the detection of more metabolites compared with MEGA-PRESS and PRESS, since it has higher resolution due to 2D acquisition. However, few studies have reported on the test–retest reliability of these MRS acquisitions for the detection of small and large brain metabolites.[8,9]

Most of the previously published studies investigating the reliability of MRS methods have focused on characterizing the test–retest reproducibility of GABA concentrations derived with MEGA-PRESS.[10–18] Almost all previous works reported coefficient of variation (CV) values only, and few studies compared intraclass correlation coefficients (ICC),[15,17–19] using mainly the MEGA-PRESS sequence. Three papers[15,17,19] discussed Pearson's correlations and only van Veenendaal et al[10] reported Bland–Altman (BA) plots. So far, two brain regions have been extensively examined: the anterior cingulate cortex (ACC) and the occipital lobe (OL). Few reliability studies investigated the posterior cingulate cortex (PCC) region and, thus, they used different sequences compared with those of the present study.[20,21] The PCC is an important hub region of the default mode network, a major resting-state brain network.[22] Its function is implicated in various disease processes including neurodegeneration[22,23] and pain,[24] and thus, a highly relevant target for detailed metabolic examination.

We focused the present investigations on GABA and Glx due to their relevance in clinical applications. GABA is the main inhibitory neurotransmitter in the brain, involved in physiological processes, ie, visual light–dark adaptation, and in several neurological and psychiatric disorders.[3] Glu, a major excitatory neurotransmitter and precursor of GABA, is the most abundant amino acid in the brain and, along with Gln, plays a crucial role in the Glu-Gln neurotransmitter cycle and in ammonia detoxification[3] and is involved in several genetic and acquired neurodegenerative processes.

In the present work, we aimed to assess the reliability of GABA and Glx measurements within the PCC in healthy participants, using PRESS, JPRESS, and MEGA-PRESS sequences applied at 3T.

Materials and Methods

Subjects

Eighteen healthy participants (median age = 29 years; range, 22–35 years; 6 males and 12 females) were recruited. The study was approved by the local Ethics Committee (Cantonal Ethics Committees of Zurich) and registered under the BASEC number 2016-00710. All participants provided written informed consent before participation in the study.

MRS Data Acquisition

All MRS measurements were performed with a 3T MR scanner (Philips Ingenia, Philips Healthcare; Best, The Netherlands) with a 32-channel head coil. The scanning protocol included a 3D anatomical T_1-weighted image with turbo-fast echo-fast inversion-recovery and gradient-echo acquisition (number of slices = 160, slice thickness = 1 mm isotropic, repetition time [TR] = 4.9 msec, echo time [TE] = 4.6 msec, field of view = 240 mm × 240 mm, flip angle = 8°, reconstruction matrix = 240 × 131 × 160). The volumetric T_1-weighted images were used to place the voxel in the PCC and to calculate the GM, WM, and CSF tissue contents within the voxel for each subject. To assure the reproducibility of the voxel placement, anatomical landmarks were used as a reference: the voxel was located laterally to the border of the splenium of corpus callosum, and above the straight sinus and tentorium cerebelli, following the shape of corpus callosum and avoiding the lateral ventricles and was always performed by the same user. The MRS data were acquired using

PRESS, JPRESS, and MEGA-PRESS sequences (Table 1 and Supplementary Table S1).

Two interleaved datasets were obtained in the MEGA-PRESS scans: in the first one, referred to as MEGA-PRESS-On, 17 msec editing pulses were applied at 1.9 ppm, whereas in the second one, called MEGA-PRESS-Off, a pulse inverted the metabolites at 7.46 ppm, which had no visual effect on the spectrum of interest between 0 and 4.7 ppm. Each reported spectrum results from the subtraction of the MEGA-PRESS-Off from the MEGA-PRESS-On data[4,10,25] and will be referred to as MEGA-PRESS, for the sake of simplicity.

JPRESS uses an echo-time series with incrementally increased echo times, providing the encoding of *J* coupling evolution in 2D, called F1 and F2, respectively. F1 is called the indirect dimension and includes the coupling information, whereas F2 contains additional coupling and chemical shift information.[26] This sequence allows for the disentangling of the signal overlap between metabolites by resolving the *J*-evolution of the signal, thereby enhancing the sensitivity.[4,8,9]

As mentioned above, chemical shift displacement effects are a challenge in MRS investigations, causing variations of the modulation patterns among the different subregions of the voxel. To reduce the chemical shift displacement artifact, we added a multipulse saturation train before each sequence, resulting in inner volume saturation (IVS) and outer volume suppression (OVS).[6,27]

For the test–retest reliability, the MRS measurements were repeated within a week (range: from 1 day to 1 week). The order of the sequences was the same for both sessions and all subjects: 1) JPRESS, 2) PRESS, and 3) MEGA-PRESS. A voxel size with extension to AP (anterior–posterior) × LR (left–right) × CC (craniocaudal) directions of the voxel of interest (VOI) = 20 × 20 × 25 mm (10 ml) for JPRESS and PRESS sequences was chosen and VOI = 30 × 30 × 30 mm (27 ml) for MEGA-PRESS. The voxel was placed in the PCC region (see Fig. 1). In all, 320 spectral averages (40 blocks × 8 averages/block) were acquired for each MEGA-PRESS sequence (TR = 1600 msec, TE = 68 msec, acquisition time = 12 min). In total, 800 spectral averages were measured for JPRESS with 100 incremented TE steps (TR = 1600 msec, initial TE = 32 msec, acquisition time = 24 min, step increment = 2 msec, and an additional water reference scan/TE step for frequency and phase correction) and 128 for PRESS (TR = 1600 msec, TE = 32 msec, acquisition time = 4 min). VAPOR (VAriable Power and Optimized Relaxation delays) pulse trains[28] were applied in all MRS sequences for optimized water suppression (see Table 1).

TABLE 1. Parameters and Analysis Pipelines for the MRS Sequences Used in This Study

Sequence	Sequence details	Analysis tool
JPRESS	TE/TR = 32/1600 msec, 800 averages, VAPOR water suppression, IVS; voxel volume = 10 mL; total scan time = 24 min	ProFit2
PRESS	TE/TR = 32/1600 msec, 128 averages, VAPOR water suppression, IVS voxel volume = 10 mL; total scan time = 4 min	LCModel
MEGA-PRESS	TE/TR = 68/1600 msec, 320 averages, editing pulses (spectral bandwidth 2 kHz) at 1.9 ppm (ON) and 7.46 ppm (OFF), VAPOR water suppression, IVS; voxel volume = 27 mL; total scan time = 12 min	LCModel

All metabolite concentrations are reported relative to the internal water concentration. The tissue composition of the voxel was determined as the composition of GM, WM, and CSF. In-house written MatLab-based scripts (MathWorks, Natick, MA) incorporating the Statistical Parametric Mapping (SPM8; Wellcome Trust Centre for Neuroimaging, London, UK) segmentation routines were used to calculate the voxel's tissue composition (WM, GM, and CSF) based on the T_1-weighted anatomical images.[29]

MRS Data Analysis

The MEGA-PRESS and PRESS data were analyzed with LCModel (Linear Combination of Model spectra) software (v. 6.3-1L), by interpreting them as a linear combination of different simulated metabolite spectra featuring a Voigt profile.[30] In the case of PRESS, 128 signals were averaged, fitted, and analyzed, whereas for MEGA-PRESS 40 blocks, ie, 320 averages, were selected. Each spectrum was fitted with a basis set, which included 18 metabolites for JPRESS and PRESS and nine metabolites for MEGA-PRESS (the subtracted spectra). In the case of PRESS, a double inversion recovery spectrum was acquired to measure the macromolecular (MM) experimental baseline. The MM baseline was included in the basis set for PRESS analyses, since there is substantial contribution of macromolecules at the applied TE = 32 msec.[31,32] The JPRESS spectroscopic data were analyzed by using ProFit2 (Prior-knowledge Fitting) including 18 metabolites.[9] The analyses of the spectra were carried out over a chemical shift range from 0.4 to 4 ppm in the case of LCModel and from 0.5 to 4 ppm in the first dimension and from –0.4 to 0.4 in the second dimension for ProFit2. Prior to data analysis, a Hankel singular value decomposition (HSVD) water filter, eddy current, and phase correction were applied.

Statistical Analysis

All statistical evaluations were performed with SPSS v. 23 (Chicago, IL). The descriptive statistics of the datasets (each with sample size n and mean $\bar{x}$) were expressed in terms of mean and standard relative error (SRE), defined according to Eq. 1:

$$SRE\,(\%) = \frac{100 \cdot SD}{\bar{x}\sqrt{n}} \qquad (1)$$

The reliability was evaluated by calculating Wilcoxon signed ranks, Pearson's adjusted *r*, ICC, CV values, and producing BA plots. The reported *P*-values were computed by assuming two-tailed

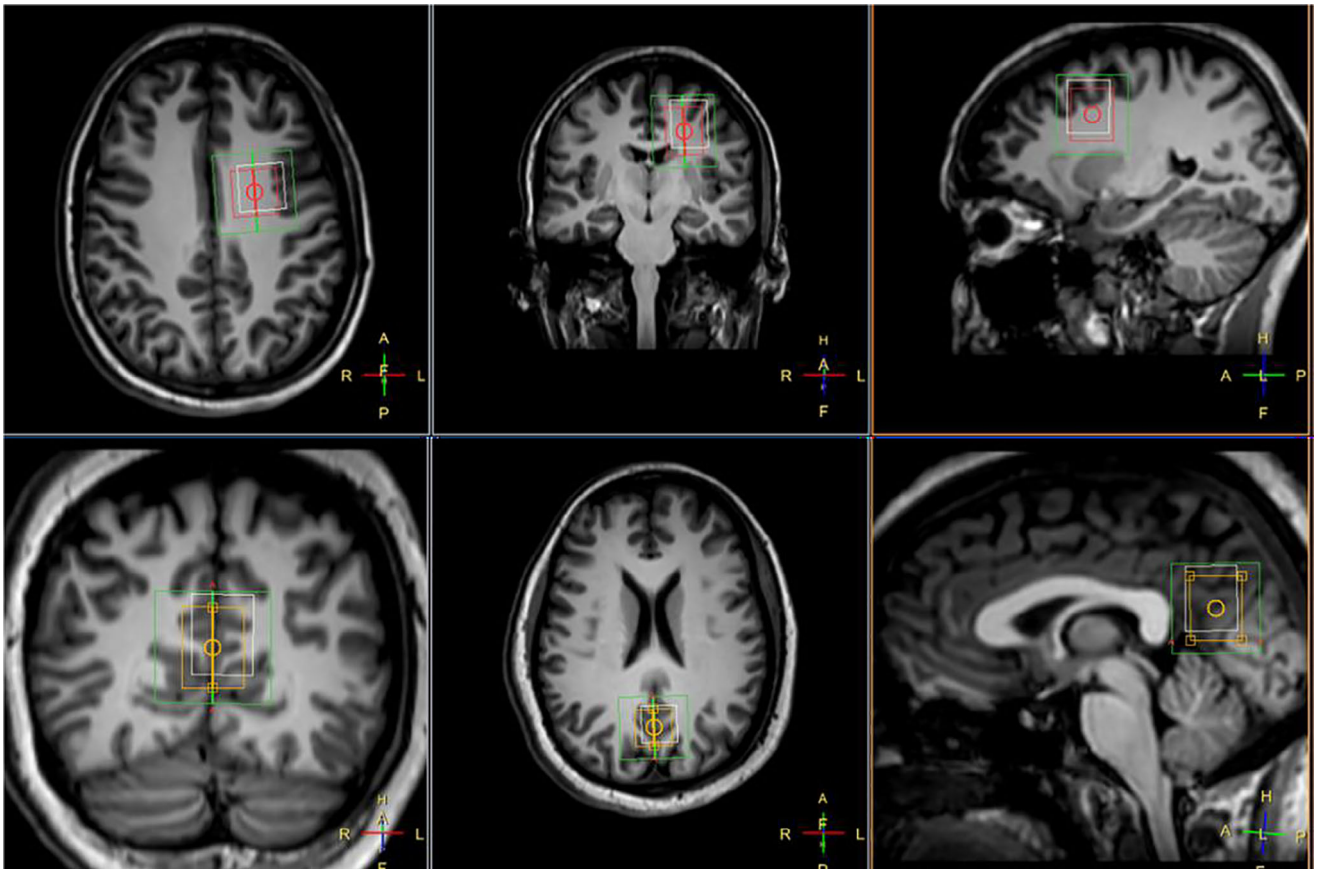

FIGURE 1: Illustration of the MRS voxel position. The MRS voxel was placed in the PCC. Shimming (green), water and lipid excited volumes at their respective frequencies/shift (white and yellow). They differ as an effect of chemical-shift displacement.

distribution. These methods aim to characterize the reliability, intended as relative scaling of the measurement error and level of agreement among paired data. The reported r_{adj} values represent the adjusted correlation coefficients, in order to use an unbiased estimator. This correction is directly applied to the r values computed by SPSS software and consists of:

$$r_{adj} = r\left[1 + \frac{1 - r^2}{2n}\right] \quad (2)$$

The CV value is defined as the ratio between the standard deviation and the means:

$$CV = \frac{\sigma}{\mu} \times 100 \quad (3)$$

The CV is the most common way to monitor this scaling, but its main limitation is that it cannot assess the relative error, when the observed values are close to zero or negative, even constraining the rank order. Another method to assess reliability that does not require a natural zero point is the evaluation of the ICC values. Like Pearson's coefficient r, ICC estimates the scaling of the measurement error in terms of correlation between two datasets. On the one hand, ICC assesses the correlation on groups, whereas r, being the ratio between the covariance and the product of the two standard deviations, focuses on the linear correlation between paired observations. Assuming a 95% confidence interval, we estimated the reliability from the obtained ICC values, according to the following convention: poor (ICC < 0.4), moderate (0.4 ≤ ICC < 0.59), good (0.6 ≤ ICC < 0.74), and excellent (ICC ≥0.75). The coefficient conveniently assesses either the absolute agreement or the consistency of the two measurements of the metabolite concentrations in the previously described test–retest (1st and 2nd measurements) or inter-sequence protocol.[33] In the case of absolute agreement, the raters' variability is also included in the formula. Intrasubject reliability ranges from zero (no reliability) to one (perfect reliability).

From a quantitative perspective, a two-way mixed model (on single measurements) for consistency between two sessions of measurements (or the two sequences) was employed to estimate the ICC values[34]:

$$\mathrm{ICC} = \frac{BMS - EMS}{BMS + (k-1)\,EMS} \quad (4)$$

where BMS = Between-measurements Mean Square variance; EMS = Mean Square Error between the measurements; k = number of repeated measurements; and n = number of subjects.

Within-subject reliability of the ROI-mean amplitudes was also estimated graphically by BA plots,[35] which track the metabolite changes of each subject across the 1st and 2nd sessions, thus enabling a visual assessment of the reliability within the subjects. The BA plots are constructed by reporting the mean of the measurements on the x-axis and the difference between the two

sessions on the *y*-axis. They serve to determine visually whether the magnitude of the differences is comparable throughout the whole range of measurements and to detect outliers. The importance of BA plots is that they assess the agreement between the datasets, since the horizontal line represents the mean of the difference set and, hence, in case of good agreement, should be simply at the zero level. A minimal error estimate of the MRS data was assessed by Cramér–Rao Lower Bounds (CRLB) for the analysis done with LCModel and by standard deviation (SD) for those with ProFit2.

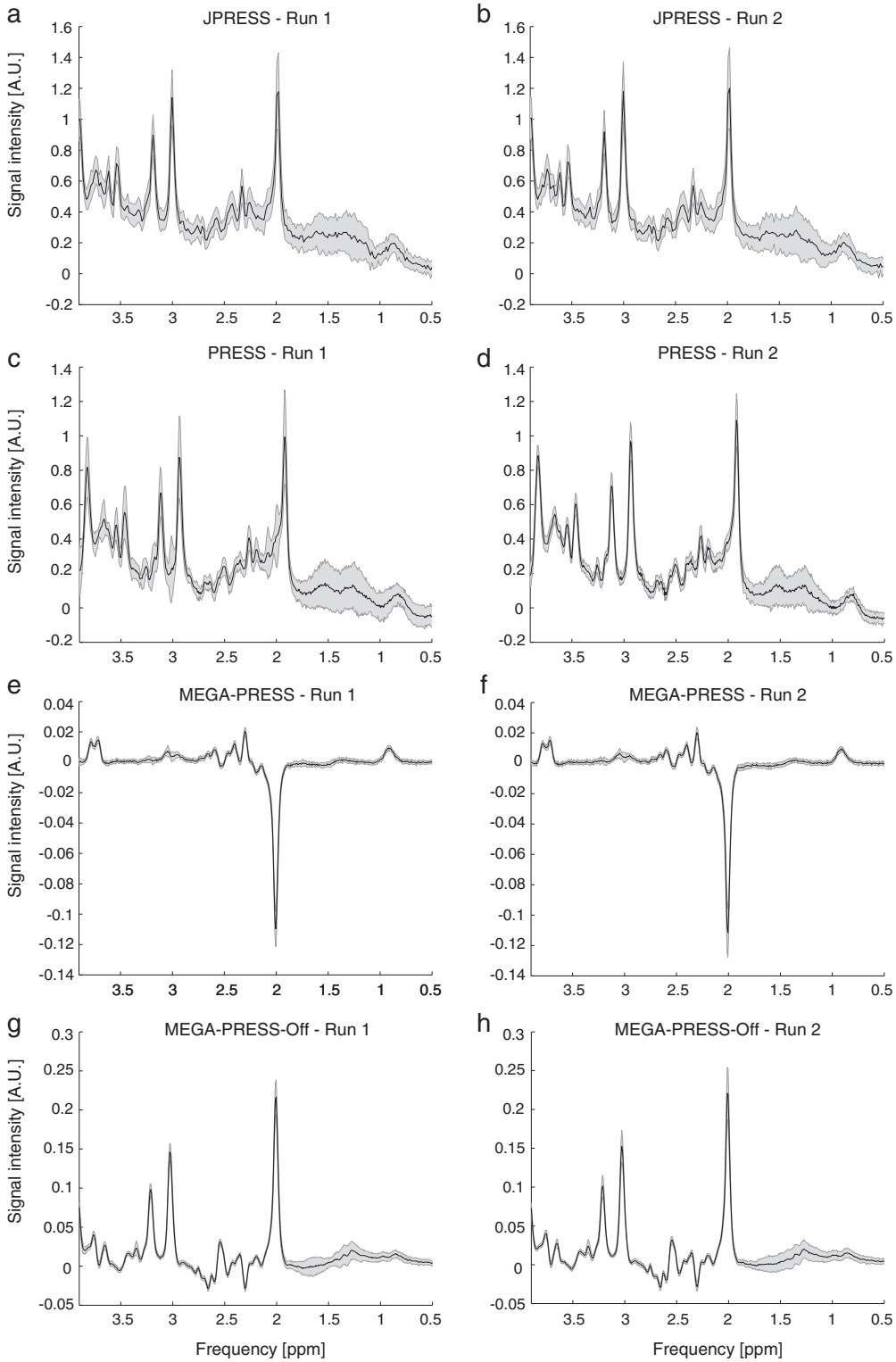

FIGURE 2: MRS spectra from the two measurements (1st and 2nd) in the voxel shown in Fig. 1. The first (a) and second (b) run of the PRESS sequence. (c) and (d) show the 1D projection of the 2D J-resolved MRS data measured with the JPRESS sequence. (e) and (f) show the MEGA-PRESS spectra (ie, MEGA-PRESS subtracted), whereas (g) and (h) the spectra of the MEGA-PRESS-Off sequence.

Results

From the T_1-weighted images used to position the voxel, 64 ± 8% of the voxel volume was occupied by GM for both the 1st and 2nd measurement datasets.

Averaged spectra are depicted in Fig. 2 for the JPRESS, PRESS, MEGA-PRESS, and MEGA-PRESS Off sequences. The mean and SD of the samples of the 1st and 2nd measurements are shown in Table 2. The mean and SD of the combined datasets (1st and 2nd measurements together) are shown in Supplementary Table S2. The first data were used to assess the test–retest reliability, whereas the second ones served for the intersequence reliability. These two assessments were accomplished by combining several statistical tools: Wilcoxon signed-rank tests (Supplementary Tables S3 and S4), Pearson's *r* correlation coefficients (Table 3, Fig. 3, Supplementary Table S5, and Supplementary Fig. S1), ICC coefficients (Table 4), and CV coefficients (Table 5) for the test–retest reliability case. The agreement between paired datasets was evaluated through BA plots (Fig. 4 for test–retest and Supplementary Fig. S2 for intersequence reliability). An example boxplot to show the test–retest reliability is presented in Fig. 5.

TABLE 2. Means, Standard Deviation (STD), Standard Error Relative to the Mean (SEM), and Standard Relative Error (SRE) for Each Sequence Employed in the Detection of the GABA and Glx (ie, Glu + Gln) Signals

GABA (*n* = 18)	MEAN	STD	SEM	SRE (%)
JPRESS (1st)	1.12	0.57	0.13	12
JPRESS (2nd)	1.48	1.01	0.24	16
PRESS (1st)	0.59	0.40	0.09	16
PRESS (2nd)	0.61	0.49	0.12	19
MEGA-PRESS (1st)	3.43	0.50	0.12	3
MEGA-PRESS (2nd)	3.21	0.50	0.12	4
MEGA-PRESS-Off (1st)	0.03	0.09	0.02	**74**
MEGA-PRESS-Off (2nd)	0.13	0.26	0.06	**48**
Glx (*n* = 18)				
JPRESS (1st)	15.54	5.50	1.30	8
JPRESS (2nd)	16.44	6.41	1.51	9
PRESS (1st)	19.35	1.76	0.41	2
PRESS (2nd)	19.11	2.00	0.47	2
MEGA-PRESS (1st)	17.41	1.91	0.45	3
MEGA-PRESS (2nd)	17.51	1.84	0.43	2
MEGA-PRESS-Off (1st)	11.00	0.93	0.22	2
MEGA-PRESS-Off (2nd)	10.84	1.41	0.33	3

High (>20%) SRE values are highlighted in bold. The parameters are estimated assuming a normally distributed population and separately for the 1st and 2nd measurements. GABA: γ-aminobutyric acid, Glu: glutamate, Gln: glutamine, Glx: Glu + Gln.

In the case of GABA, a moderate correlation was observed only in the case of MEGA-PRESS (r = 0.54, ICC = 0.5, CV = 8.8%). Lower correlation coefficients were observed for JPRESS (r = 0.39, ICC = 0.3, CV = 49.7%) and PRESS (r = 0.31, ICC = 0.3, CV = 63.3%), but the BA plots show an acceptable agreement for all three sequences. For the intersequence reliability results, the combined datasets exhibit remarkably different mean values of GABA for JPRESS (1.30 ± 0.14, SRE = 11%), PRESS (0.60 ± 0.07, SRE = 12%), and MEGA-PRESS (3.32 ± 0.08, SRE = 3%). The Wilcoxon signed-rank test revealed highly significant ($P < 0.001$) differences between GABA values derived with all sequences. However, the low mean concentrations observed with PRESS should be evaluated with caution, as these values are affected by numerous zeros in the concentration values and high (>100%) relative CRLB in several subjects.

Regarding Glx, comparable concentration values were obtained with JPRESS (15.99 ± 0.98, SRE = 6%), PRESS (19.23 ± 0.31, SRE = 2%), and MEGA-PRESS (17.46 ± 0.31, SRE = 2%). The good test–retest reliability for Glx in all sequences is reflected by the absence of statistically significant differences, as shown by Wilcoxon signed-rank tests and by BA plots. Moderate to excellent correlations between sessions were achieved with PRESS (r = 0.87, ICC = 0.9, CV = 2.9%) and MEGA-PRESS (r = 0.70, ICC = 0.7, CV = 5.3%), but a lower precision was observed for JPRESS (r = 0.39, ICC = 0.4, CV = 20.1%). The intersequence Wilcoxon signed-rank tests and BA plots indicate that there are statistically significant differences between PRESS and JPRESS [P(Wilcoxon) = 0.004], and PRESS and MEGA-PRESS [P(Wilcoxon) < 0.001] and a trend towards a significant difference in GABA levels measured with JPRESS and MEGA-PRESS sequences [P(Wilcoxon) = 0.08]. Interestingly, there is no intersequence correlation, except for the moderate correlation in GABA values estimated with PRESS and MEGA-PRESS (r = 0.59). Finally, the Wilcoxon signed-rank test shows that the best test–retest agreement for distinguishing Glu from Gln is achieved with JPRESS (Supplementary Table S3).

TABLE 3. Pearson's *r* Correlation Adjusted Coefficients Relative to Two Scan Timepoints (1st and 2nd Measurements) for JPRESS, PRESS, and MEGA-PRESS (*n* = 18)

		JPRESS	PRESS	MEGA-PRESS	MEGA-PRESS-Off
GABA	*r*	0.39	0.31	**0.54***	0.05
	P (2-tailed)	0.11	0.23	0.02	0.85
Gln	*r*	**0.61****	**0.48***	0.38	0.30
	P (2-tailed)	0.009	0.05	0.13	0.24
Glu	*r*	0.33	**0.77****	**0.58***	**0.72****
	P (2-tailed)	0.20	< 0.001	0.01	0.001
Glx	*r*	0.39	**0.87****	**0.70****	**0.73****
	P (2-tailed)	0.12	< 0.001	0.002	0.001

**$P < 0.01$, *$P < 0.05$.
GABA: γ-aminobutyric acid, Glu: glutamate, Gln: glutamine, Glx: Glu + Gln.

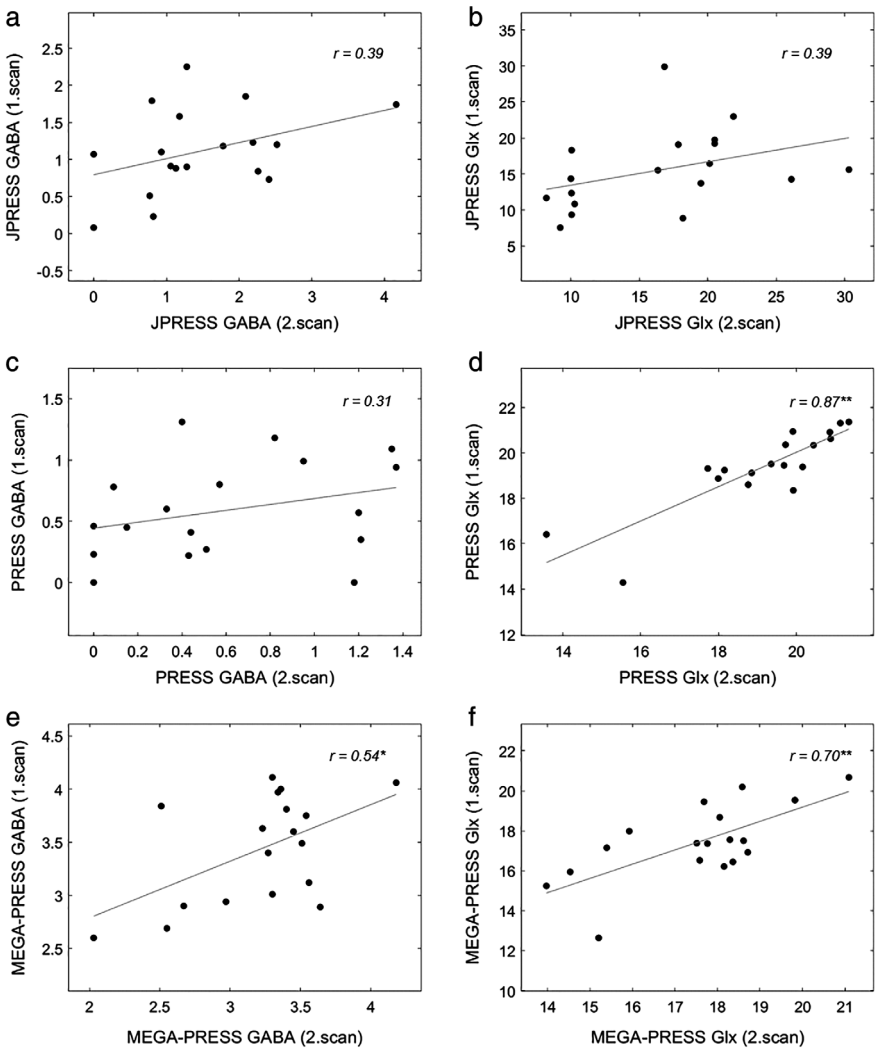

FIGURE 3: Pearson's *r* correlation plots for GABA (a,c,e) and Glx (b,d,f) for each of the three MRS sequences to compare the data acquired in the two different scan timepoints (1st and 2nd measurement). From top to bottom: JPRESS, PRESS, and MEGA-PRESS. *$P \leq 0.05$, **$P \leq 0.01$.

TABLE 4. Intraclass Correlation Coefficients (ICC) Relative to Two Scan Timepoints (1st and 2nd Measurements) for JPRESS, PRESS, and MEGA-PRESS (*n* = 18)

	JPRESS	PRESS	MEGA-PRESS	MEGA-PRESS-Off
GABA	0.3	0.3	**0.5**	< 0.1
Gln	**0.6**	**0.5**	**0.4**	0.3
Glu	0.3	**0.8**	**0.6**	**0.7**
Glx	**0.4**	**0.9**	**0.7**	**0.7**

GABA: γ-aminobutyric acid, Glu: glutamate, Gln: glutamine, Glx: Glu + Gln. ICC classification: poor (ICC < 0.4), fair (0.4 ≤ ICC < 0.59), good (0.6 ≤ ICC < 0.74), and excellent (ICC ≥0.75). Fair and good correlation are written in bold, whereas excellent correlations are underlined bold. Method of estimate: two-way mixed consistency on single measures ICC.

TABLE 5. CV Values Relative to Difference Between Two Scan Timepoints (1st and 2nd Measurements) for JPRESS, PRESS, and MEGA-PRESS (*n* = 18)

	JPRESS	PRESS	MEGA-PRESS	MEGA-PRESS-Off
GABA	49.7%	63.3%	**8.8%**	110.3%
Gln	34.2%	**9.8%**	47.0%	97.2%
Glu	22.2%	**4.5 %**	**6.1%**	**4.2%**
Glx	**20.1%**	**2.9%**	**5.3 %**	**5.1%**

A lower CV value reflects a higher test-retest reliability. We assume that CV values below the threshold of 20% are reliable. GABA: γ-aminobutyric acid, Glu: glutamate, Gln: glutamine, Glx: Glu + Gln.

Discussion

Most previous studies evaluated the MEGA-PRESS sequence for quantification of GABA, Glu, Gln, and Glx or GABA only.[10–18] Two studies quantified Glx by JPRESS.[36,37] CV evaluation for test–retest reliability has been applied extensively to all sequences in several investigations, as for MEGA-PRESS[16] and JPRESS.[36] Some ICC analyses were already reported.[15,17,18] Only two previous studies[15,17] applied the Pearson's *r* correlation method to measure the reliability of GABA detection by MEGA-PRESS and one[19] to measure the reliability of Glx detection by PRESS. Moreover, only two investigations,[20,21] one using semi-LASER (Localized by Adiabatic SElective Refocusing) and another using 2D L-COSY (Localized COrrelational SpectroscopY) sequences at 3T and 7T, reported on the PCC region of the brain, despite its important function in brain regulation and homeostasis and its extensive implication in several neurodegenerative processes.

To minimize statistical overinterpretation due to the limited size of the sample, we assessed our data by combining mutually unbiased methods, such as Wilcoxon signed-rank tests, BA plots, and calculations of Pearson's *r* correlation, ICC, and CV coefficients.

The advantage in terms of test–retest reliability of edited sequences like MEGA-PRESS in the detection of GABA is confirmed by the higher precision, as reported in this study. Despite the absence of statistically significant differences in the Wilcoxon signed-rank test for test–retest reliability, PRESS seems to be unsuitable for GABA detection. This is not only reflected by the lowest correlation *r* and ICC but also by the high CRLB values (>100%) and frequent presence of zeroes, indicating that PRESS data show low sensitivity to the GABA peak, which overlaps with signals from other metabolites. Using JPRESS, the GABA signal can be clearly resolved. Despite the lower test–retest reliability in comparison to MEGA-PRESS, JPRESS may still be the most accurate among the employed sequences; however, it also requires the longest acquisition time. In fact, MEGA-PRESS data may overestimate GABA concentrations, due to the limited flexibility in the baseline definition and, hence, higher risk to compute values affected by MM contamination.[38] Alternatively, the overestimate of GABA from MEGA-PRESS may arise from differences in scaling factors within the basis set; an unequivocal solution would be to refer the observation to a phantom that can be assumed as the gold standard, for assessing the absolute accuracy in the concentration estimate,

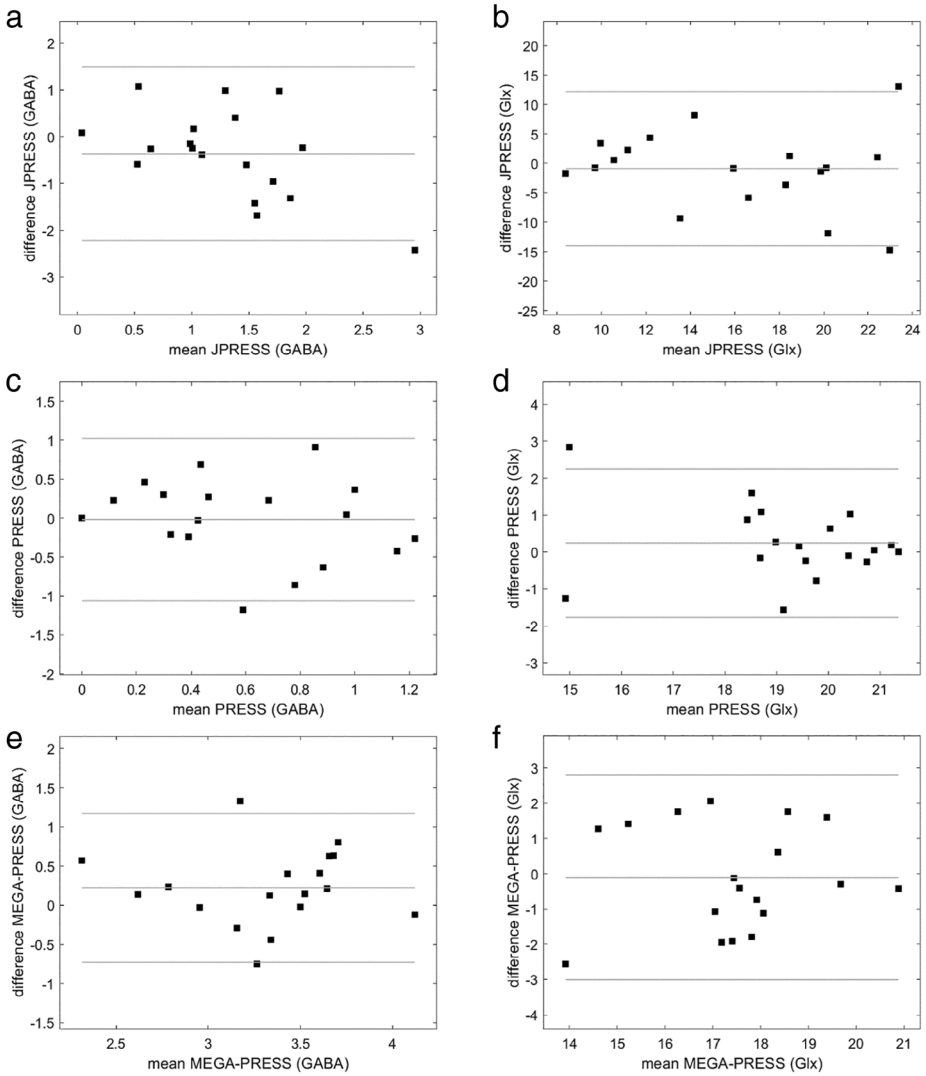

FIGURE 4: Bland–Altman plots for GABA (a,c,e) and Glx (b,d,f) for the three MRS sequences showing the agreement between the two different scan timepoints (1st and 2nd measurement). The central gray line indicates the mean value of the difference dataset (μ) and the upper and lower ones are the limits μ + 1.96SD and μ-1.96SD, respectively. From top to bottom: JPRESS, PRESS, and MEGA-PRESS.

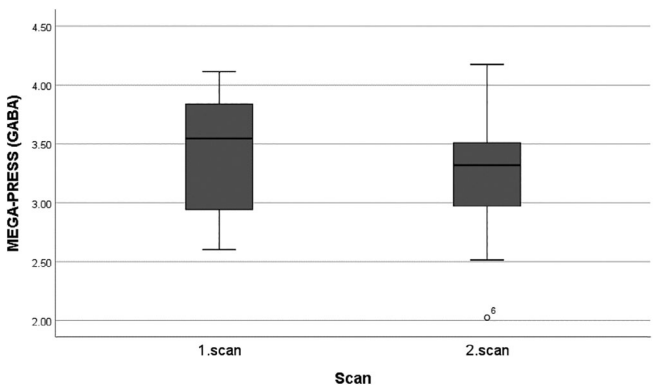

FIGURE 5: Conventional boxplot showing the concentration distribution of GABA acquired with MEGA-PRESS in the first (left) and second measurements (right). The box indicates the interquartile range (IQR, MEAN ± 0.6745 × SD) and the upper and lower whiskers denote the limit (MEAN ± 2.698 × SD interval). Outliers are marked with a single dot.

but this is beyond the scope of this study. Previous studies have reported correlation coefficients in accordance with those derived in the present study, ie, r = 0.54 for MEGA-PRESS.[17] The CV value we obtained with MEGA-PRESS (8.8%) is also comparable to the typical values reported in the literature, which vary between 6% and 8%.[11,12,16] Lower values were reported by Near et al[17] (4.3% in the OL) and Shungu et al,[15] who obtained values for the dorsolateral prefrontal cortex (DLPFC) of 4% (using a single-channel coil) and 2% (using an eight-channel coil). In the case of MEGA-PRESS investigations performed at 7T, similar values were reported in the ACC, whereas different results at the same field strength were obtained using STEAM (STimulated Echo Acquisition Mode) and semi-LASER sequences, the latter exhibiting the best reliability for Glu (3%). Compared with MEGA-PRESS, semi-LASER proved to be less reliable in detecting GABA (CV values between 13% to 22%).[11,20] However, CV values in

the detection of GABA are often claimed to be dependent on several factors (e.g., the ROI), and therefore limit the possibility for meaningful comparisons.[16,36]

In general, Glx estimations are comparable for all three sequences, with a minimal variation reported for JPRESS, followed by MEGA-PRESS and PRESS. The Wilcoxon signed-rank tests did not detect statistically significant differences between the 1st and 2nd measurements, with the exception of Gln by MEGA-PRESS. The highest precision is exhibited by PRESS for Glx, Glu, and Gln. In addition, the *r* correlation coefficients confirm the high precision achieved by PRESS for Glx (0.87) and Glu (0.77), which are comparable to, or slightly better than, values reported in the literature[19] (0.64 and 0.79, respectively), although the latter values were obtained with a smaller voxel size (3.4 mL). Indeed, PRESS is a conventional and reliable sequence with a favorably short acquisition time, as indicated by our results and previous works. Its limitation becomes evident in the case of low concentration metabolites like GABA or when two metabolites with similar overlapping spectra need to be detected separately, as, eg, in the case of Glu and Gln. Moreover, with MEGA-PRESS we could detect Glx with CV = 5.3% and confirm the good agreement, as already reported in a previous work with a smaller sample size.[10] MEGA-PRESS and JPRESS are less precise than PRESS (showing higher CV values), but all Glx concentration values resulted in fairly comparable results. MEGA-PRESS is generally more prone to motion of the subject and hardware instabilities than JPRESS, which increases the risk of subtraction artifacts,[39] although the acquisition of interleaved edit ON and OFF data, combined with preprocessing methods like frequency and phase correction and outlier rejection, can improve reproducibility and reduce the likelihood of subtraction artifacts.[12,40] The reliability of JPRESS is dependent on using different acquisitions with incremented TE, which allows for a more precise separation of the contributions from Glu and Gln. TE series measurements allow for the disentangling of the different modulations of the two metabolites, clearly enhancing the reliability; however, at the price of longer acquisition times.[8] Our reliability parameters are in good agreement with those reported in the literature. Particularly, the reported CV values are comparable to those obtained with PRESS and MEGA-PRESS.[10] Regarding JPRESS, we obtained a slightly higher CV = 20.1%, as compared with 10% and 16% from previous reports.[36,37] There is no evident explanation for the lower precision of JPRESS in comparison to the other sequences.

A trend towards consistency in Glx concentrations was obtained only between the JPRESS and MEGA-PRESS sequences [*P*(Wilcoxon) = 0.08]. The significant differences in Glx derived with the other intersequence tests are most likely due to the high test–retest precision, which makes the intersequence reliability tests extremely strict in terms of error tolerance. Furthermore, statistical error in the hypothesis tests do not take into consideration systematic error due to other factors, which may play a major role, such as the use of different postprocessing protocols for the spectral fitting routines and of different metabolite basis sets.

A major limitation of the study was that, due to differences in voxel volume and scan time between the sequences, the relative signal-to-noise ratio (SNR) was not identical, leading to differences in variability between measurements, which may affect the assessment of reproducibility. However, given the relatively short scan time for the PRESS acquisition in comparison with the other sequences, the apparently higher precision of PRESS for Glx in particular is unlikely to be due to differences in SNR between the sequences. Moreover, it appears to reflect a genuine benefit in terms of precision. An advantage of the protocols used in the present study is that, since each sequence was tested in its standard protocol implementation, the variability values measured should reflect those associated with a typical protocol, as used in many previous and ongoing studies. Furthermore, the present study did not assess the precision of other MRS localization and acquisition methods, used in the detection of GABA and Glx, such as semi-LASER, STEAM, and 2D COSY.[10,11,20] The absolute estimate of metabolite concentrations (assessed by the accuracy) would require measurements from a phantom to be taken as a gold standard. Some practical limitations have to be taken into account, like the limited sample size (n = 18), the use of a larger voxel size in the case of MEGA-PRESS to achieve a sufficient SNR, and the different postprocessing tools used to analyze the PRESS and MEGA-PRESS data with LCModel and the JPRESS data with Profit2.

In conclusion, our study presents a somewhat rigorous statistical evaluation of the test–retest and intersequence reliability for the JPRESS, PRESS, and MEGA-PRESS sequences, used for measuring the concentration of GABA and Glx (Glu + Gln) metabolites in the PCC. We show that edited-pulse sequences are suitable for the detection of GABA with 30% precision. Nevertheless, MEGA-PRESS data tend to overestimate the GABA concentration due to MM contamination and JPRESS may be a more accurate alternative, despite the lower test–retest reliability and longer scan time.

In the case of Glx (Glu + Gln), all three sequences exhibited moderate to excellent levels of test–retest reliability in terms of correlation and agreement. Some suggestions are also provided to help choosing the most appropriate sequences, according to the type of investigation, the metabolite of interest (eg, GABA or Glx (Glu + Gln)), the clinical setup (eg, regarding the availability of certain sequences), and time constraints.

Acknowledgment

Contract grant sponsor: Saudi Arabian Cultural Mission in Paris, France.

We thank all the healthy volunteers who participated in this study.

References

1. Henry ME, Lauriat TL, Shanahan M, Renshaw PF, Jensen JE. Accuracy and stability of measuring GABA, glutamate, and glutamine by proton

magnetic resonance spectroscopy: A phantom study at 4 Tesla. Journal of magnetic resonance. 2011;208:210–218.

2. Harris AD, Saleh MG, Edden RA. Edited 1H magnetic resonance spectroscopy in vivo: Methods and metabolites. Magn Reson Med 2017;77: 1377–1389.
3. De Graaf RA. *In vivo NMR spectroscopy: Principles and techniques*. Hoboken, NJ: John Wiley & Sons; 2013.
4. Puts NA, Edden RA. In vivo magnetic resonance spectroscopy of GABA: A methodological review. Prog Nucl Magn Reson Spectrosc 2012;60:29.
5. Tedeschi G, Bertolino A, Righini A, et al. Brain regional distribution pattern of metabolite signal intensities in young adults by proton magnetic resonance spectroscopic imaging. Neurology 1995;45:1384–1391.
6. Edden RA, Barker PB. Spatial effects in the detection of γ-aminobutyric acid: Improved sensitivity at high fields using inner volume saturation. Magn Reson Med 2007;58:1276–1282.
7. Henning A, Schär M, Kollias SS, Boesiger P, Dydak U. Quantitative magnetic resonance spectroscopy in the entire human cervical spinal cord and beyond at 3T. Magn Reson Med 2008;59:1250–1258.
8. Schulte RF, Lange T, Beck J, Meier D, Boesiger P. Improved two-dimensional J-resolved spectroscopy. NMR Biomed 2006;19:264–270.
9. Fuchs A, Boesiger P, Schulte RF, Henning A. ProFit revisited. Magn Reson Med 2014;71:458–468.
10. van Veenendaal TM, Backes WH, van Bussel FCG, et al. Glutamate quantification by PRESS or MEGA-PRESS: Validation, repeatability, and concordance. Magn Reson Imaging 2018;48:107–114.
11. Wijtenburg SA, Rowland LM, Edden RA, Barker PB. Reproducibility of brain spectroscopy at 7T using conventional localization and spectral editing techniques. J Magn Reson Imaging 2013;38:460–467.
12. Brix MK, Ersland L, Hugdahl K, et al. Within-and between-session reproducibility of GABA measurements with MR spectroscopy. J Magn Reson Imaging 2017;46:421–430.
13. Bogner W, Gruber S, Doelken M, et al. In vivo quantification of intracerebral GABA by single-voxel 1H-MRS—How reproducible are the results? Eur J Radiol 2010;73:526–531.
14. Long Z, Dyke JP, Ma R, Huang CC, Louis ED, Dydak U. Reproducibility and effect of tissue composition on cerebellar γ-aminobutyric acid (GABA) MRS in an elderly population. NMR Biomed 2015;28: 1315–1323.
15. Shungu DC, Mao X, Gonzales R, et al. Brain GABA detection in vivo with the J-editing 1H MRS technique: A comprehensive methodological evaluation of sensitivity enhancement, macromolecule contamination and test-retest reliability. NMR Biomed 2016;29:932.
16. O'Gorman RL, Michels L, Edden RA, Murdoch JB, Martin E. In vivo detection of GABA and glutamate with MEGA-PRESS: Reproducibility and gender effects. J Magn Reson Imaging 2011;33:1262–1267.
17. Near J, Ho Y-CL, Sandberg K, Kumaragamage C, Blicher JU. Long-term reproducibility of GABA magnetic resonance spectroscopy. Neuroimage 2014;99:191–196.
18. Geramita M, van der Veen JW, Barnett AS, et al. Reproducibility of prefrontal γ-aminobutyric acid measurements with J-edited spectroscopy. NMR Biomed 2011;24:1089–1098.
19. Liu X-L, Li L, Li J-N, Rong J-H, Liu B, Hu Z-X. Reliability of glutamate quantification in human nucleus accumbens using proton magnetic resonance spectroscopy at a 70-cm wide-bore clinical 3T MRI system. Front Neurosci 2017;11:686.
20. Terpstra M, Cheong I, Lyu T, et al. Test-retest reproducibility of neurochemical profiles with short-echo, single-voxel MR spectroscopy at 3T and 7T. Magn Reson Med 2016;76:1083–1091.
21. Van de Bank B, Emir U, Boer V, et al. Multi-center reproducibility of neurochemical profiles in the human brain at 7 T. NMR Biomed 2015; 28:306–316.
22. Leech R, Sharp DJ. The role of the posterior cingulate cortex in cognition and disease. Brain 2013;137:12–32.
23. Riese F, Gietl A, Zölch N, et al. Posterior cingulate γ-aminobutyric acid and glutamate/glutamine are reduced in amnestic mild cognitive impairment and are unrelated to amyloid deposition and apolipoprotein E genotype. Neurobiol Aging 2015;36:53–59.
24. Aguila M-ER, Rebbeck T, Leaver AM, et al. The association between clinical characteristics of migraine and brain GABA levels: An exploratory study. J Pain 2016;17:1058–1067.
25. Mescher M, Merkle H, Kirsch J, Garwood M, Gruetter R. Simultaneous in vivo spectral editing and water suppression. NMR Biomed 1998;11: 266–272.
26. Ryner LN, Sorenson JA, Thomas MA. Localized 2D J-resolved 1H MR spectroscopy: Strong coupling effects in vitro and in vivo. Magn Reson Imaging 1995;13:853–869.
27. Edden RA, Schar M, Hillis AE, Barker PB. Optimized detection of lactate at high fields using inner volume saturation. Magn Reson Med 2006;56:912–917.
28. Tkáè I, Starèuk Z, Choi IY, Gruetter R. In vivo 1H NMR spectroscopy of rat brain at 1 ms echo time. Magn Reson Med 1999;41:649–656.
29. Gasparovic C, Song T, Devier D, et al. Use of tissue water as a concentration reference for proton spectroscopic imaging. Magn Reson Med 2006;55:1219–1226.
30. Provencher SW. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. Magn Reson Med 1993;30:672–679.
31. Provencher S. LCModel & LCMgui User's Manual. LCModel version 6.3-1L. 2016.
32. Giapitzakis IA, Borbath T, Murali-Manohar S, Avdievich N, Henning A. Investigation of the influence of macromolecules and spline baseline in the fitting model of human brain spectra at 9.4 T. Magn Reson Med 2018 [Epub ahead of print].
33. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1:30.
34. Koo T, Li M. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15: 155–163.
35. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1: 307–310.
36. Prescot AP, Renshaw PF. Two-dimensional J-resolved proton MR spectroscopy and prior knowledge fitting (ProFit) in the frontal and parietal lobes of healthy volunteers: Assessment of metabolite discrimination and general reproducibility. J Magn Reson Imaging 2013;37:642–651.
37. Mullins PG, Chen H, Xu J, Caprihan A, Gasparovic C. Comparative reliability of proton spectroscopy techniques designed to improve detection of J-coupled metabolites. Magn Reson Med 2008;60:964–969.
38. Edden RA, Puts NA, Barker PB. Macromolecule-suppressed GABA-edited magnetic resonance spectroscopy at 3T. Magn Reson Med 2012;68:657–661.
39. Yahya A. Metabolite detection by proton magnetic resonance spectroscopy using PRESS. Prog Nucl Magn Reson Spectrosc 2009;3:183–198.
40. Edden RA, Puts NA, Harris AD, Barker PB, Evans CJ. Gannet: A batch-processing tool for the quantitative analysis of gamma-aminobutyric acid-edited MR spectroscopy spectra. J Magn Reson Imaging 2014;40: 1445–1452.