Received: 16 July 2020 | Revised: 22 September 2020 | Accepted: 17 October 2020

DOI: 10.1002/mrm.28587

NOTE

Magnetic Resonance in Medicine

Repeatability and reliability of GABA measurements with magnetic resonance spectroscopy in healthy young adults

Jessica M. Duda[1] | **Amelia D. Moser**[1] | **Chun S. Zuo**[2,3] | **Fei Du**[2,3,4] | **Xi Chen**[2,3,4] | **Sarah Perlo**[1] | **Christine E. Richards**[1] | **Nara Nascimento**[1] | **Maria Ironside**[1,3] | **David J. Crowley**[1] | **Laura M. Holsen**[3,5,6] | **Madhusmita Misra**[3,7] | **James I. Hudson**[3,8] | **Jill M. Goldstein**[3,9,10] | **Diego A. Pizzagalli**[1,2,3]

[1]Center for Depression, Anxiety and Stress Research, McLean Hospital, Belmont, Massachusetts, USA

[2]McLean Imaging Center, McLean Hospital, Belmont, Massachusetts, USA

[3]Harvard Medical School, Boston, Massachusetts, USA

[4]Schizophrenia and Bipolar Research Program, McLean Hospital, Belmont, Massachusetts, USA

[5]Department of Medicine, Division of Women's Health, Brigham & Women's Hospital, Boston, Massachusetts, USA

[6]Department of Psychiatry, Brigham & Women's Hospital, Boston, Massachusetts, USA

[7]Department of Pediatrics, Division of Pediatric Endocrinology, Massachusetts General Hospital, Boston, Massachusetts, USA

[8]Biological Psychiatry Laboratory, McLean Hospital, Belmont, Massachusetts, USA

[9]Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, USA

[10]Department of Obstetrics and Gynecology, Massachusetts General Hospital, Boston, Massachusetts, USA

Correspondence
Diego A. Pizzagalli, Center for Depression, Anxiety and Stress Research, McLean Hospital, 115 Mill Street, Belmont, MA 02478, USA.
Email: dap@mclean.harvard.edu

Funding information
National Institute of Mental Health (Grant Nos. R01MH108602 and R37MH068376)

Purpose: Gamma-aminobutyric acid (GABA) abnormalities have been implicated in a range of neuropsychiatric disorders. Despite substantial interest in probing GABA in vivo, human imaging studies relying on magnetic resonance spectroscopy (MRS) have generally been hindered by technical challenges, including GABA's relatively low concentration and spectral overlap with other metabolites. Although past studies have shown moderate-to-strong test-retest repeatability and reliability of GABA within certain brain regions, many of these studies have been limited by small sample sizes.

Methods: GABA+ (macromolecular-contaminated) test-retest reliability and repeatability were assessed via a Meshcher-Garwood point resolved spectroscopy (MEGA-PRESS) MRS sequence in the rostral anterior cingulate cortex (rACC; $n = 21$) and dorsolateral prefrontal cortex (dlPFC; $n = 20$) in healthy young adults. Data were collected on a 3T scanner (Siemens Prisma, Siemens Healthcare, Erlangen, Germany) and GABA+ results were reported in reference to both total creatine (GABA+/tCr) and water (GABA+/water).

Results: Results showed strong test-retest repeatability (mean GABA+/tCr coefficient of variation [CV] = 4.6%; mean GABA+/water CV = 4.0%) and reliability (GABA+/tCr intraclass correlation coefficient [ICC] = 0.77; GABA+/water ICC =

Jessica M. Duda and Amelia D. Moser contributed equally to this work.

0.87) in the dlPFC. The rACC showed acceptable (but comparatively lower) repeatability (mean GABA+/tCr CV = 8.0%; mean GABA+/water CV = 7.5%), yet low-moderate reliability (GABA+/tCr ICC = 0.40; GABA+/water ICC = 0.44).

Conclusion: The present study found excellent GABA+ MRS repeatability and reliability in the dlPFC. The rACC showed inferior results, possibly because of a combination of shimming impedance and measurement error. These data suggest that MEGA-PRESS can be utilized to reliably distinguish participants based on dlPFC GABA+ levels, whereas the mixed results in the rACC merit further investigation.

KEYWORDS

anterior cingulate, dorsolateral prefrontal cortex, gamma-aminobutyric acid, LCModel, MEGA-PRESS, magnetic resonance spectroscopy

1 | INTRODUCTION

Gamma-aminobutyric acid (GABA) has been implicated in the pathogenesis of a range of psychiatric disorders including major depressive disorder (MDD) and schizophrenia.[1-4] Prior studies using magnetic resonance spectroscopy (MRS) have identified decreased GABA in the occipital cortex[4] (OCC), anterior cingulate cortex[2] (ACC), and dorsolateral prefrontal cortex[3] (dlPFC) in individuals with MDD, in line with reports of reductions in GABAergic interneurons in the dlPFC emerging from postmortem studies.[5] Likewise, postmortem and animal studies indicate possible GABAergic dysfunction in schizophrenia, although in vivo studies remain equivocal.[1,4,6] Despite evidence of GABAergic abnormalities across these disorders, the precise mechanisms of GABA dysfunction remain unclear. Therefore, continued investigation of GABA in vivo is necessary to elucidate its possible contribution to the etiology and maintenance of psychiatric illness.

MRS has become widely used for GABA quantification; its noninvasive design—paired with recent technological advances—make it a particularly attractive option.[4,7] Nevertheless, MRS quantification of GABA has been challenged by the metabolite's relatively low concentration, considerable spectral overlap with other metabolites, and macromolecular contamination.[7,8] Given these difficulties, studies testing the reliability and agreement of MRS-derived GABA levels are critical to ensuring appropriate interpretation of anomalies that may be found in the context of pathology.

As summarized in Table 1, past research has shown moderate-to-strong test-retest repeatability and reliability of GABA MRS in the cingulate cortex and OCC.[8-21] Several additional studies have examined other areas, including motor regions,[21-24] Broca's area,[9] and the dlPFC,[22,25,26] among others.[27-29] However, many studies have been limited by small sample sizes. Accordingly, replication studies are necessary, particularly as the homogeneity of the magnetic field has been shown to vary with between-subjects factors such as sinus shape and size, thereby impacting shimming and MRS data quality.[30,31]

MRS test-retest studies have also varied in their use of statistical methods and terminology. For example, many past studies used the terms "repeatability," "reliability," "reproducibility," and "agreement" seemingly interchangeably, leading to confusion about the exact constructs and theoretical underpinnings in question. In the present study, we use "repeatability" and "agreement" to refer to the degree of similarity between multiple measurements taken from a subject under identical conditions.[32] In contrast, we use "reliability" to refer to the capacity of the methodology to distinguish among subjects.[32,33] Notably, many test-retest studies have reported both repeatability metrics such as coefficients of variation (CVs) and reliability metrics such as intraclass correlation coefficients (ICCs) with minimal discussion of the fundamental differences between these types of measures. We believe further clarity regarding the respective implications of each metric is needed within the MRS literature.

In this context, the main goal of the present study was to evaluate rostral ACC (rACC) and dlPFC GABA test-retest repeatability and reliability with MRS in a larger sample of healthy young adults, with careful distinction and discussion of the agreement and reliability of GABA measurements.

2 | METHODS

2.1 | Sample

Participants ($N = 35$) were control subjects with no history of psychopathology drawn from a larger study investigating sex differences and stress in young adults with MDD. Absence of current or past psychopathology was ascertained using a semistructured clinical interview (Structured Clinical Interview for the DSM-5 [SCID-5]),[34] which was performed by a doctoral- or master's-level clinician. All participants

TABLE 1 Prior GABA MEGA-PRESS MRS repeatability and reliability studies at 3T

Reference	Brain region	*N*	Statistical methodology	Cr-referenced GABA findings	Water-referenced GABA findings
Baeshen et al, 2020[8]	PCC	18	Wilcoxon signed-rank tests	—	$Z = -1.85$
			Pearson's *r*		
			ICC		$r = 0.54$
			CV		ICC = 0.5
			BA Plots		CV = 8.8%
Ferland et al, 2019[24]	Left sensorimotor cortex (SMC)	10	Pearson's *r*	$r = 0.65$	$r = 0.82$
			ICC	ICC = 0.67	ICC = 0.81
			CV	CV = 14%	CV = 10%
Mikkelsen et al, 2019[27]	Medial parietal lobe	284[a]	CV	—	GABA+ CV = 16.9%
					GABA' CV = 28.8%
Brix et al, 2017[9]	ACC	21	CV	—	ACC: CV = 6%–14%
	Left Broca's area				Broca's area: CV = 4%–7%
Mikkelsen et al, 2017[28]	Medial parietal lobe	272[a]	CV	GABA+ CV = 12.0%	—
				GABA' CV = 27.6%	
Yasen et al, 2017[22]	dlPFC	13	ICC	dlPFC: ICC = 0.97; CV = 27.9%	—
	Primary motor cortex		CV	Motor: ICC = 0.93; CV = 21.0%	
Greenhouse et al, 2016[21]	Lateral PFC (lPFC)	28	Pearson's *r*	lPFC: $r = 0.75$, CV = 4.6%	—
	SMC			SMC: $r = 0.64$, CV = 3.9%	
	Dorsal premotor cortex (dPMC)		CV	dPMC: $r = 0.63$, CV = 3.9%	
	OCC			OCC: $r = 0.52$, CV = 5.3%	
Mikkelsen et al, 2016[10]	OCC	15	ICC	—	OCC: GABA+ ICC = 0.67; CV = 4.0%
					OCC: GABA' ICC = 0.72; CV = 8.6%
	ACC		CV		ACC: GABA+ ICC = 0.16; CV = 14.8%
					ACC: GABA' ICC = 0.41; CV = 12.6%
Shungu et al, 2016[25]	dlPFC	6	Pearson's *r*	$r = 0.98$	$r = 0.98$
			ICC	ICC = 0.97	ICC = 0.98
			CV	CV = 1.72%	CV = 1.25%
Long et al, 2015[29]	Cerebellar dentate	5	CV	Left, corrected: CV = 6.1%–13.4%	Left, corrected: CV = 5.3%–11.9%
				Right, corrected: CV = 5.0%–8.0%	Right, corrected: CV = 5.0%–8.5%
Gaetz et al, 2014[23]	ROIs in these areas:	5	CV		—
	Motor			Motor: CV = 9%	
	Visual			Visual: CV = 11%	
	Auditory			Auditory: CV = 9%	
Near et al, 2014[13]	OCC	17	Pearson's *r*	$r = 0.53$	—
			ICC	ICC = 0.52	
			CV	CV = 4.3%	

(Continues)

TABLE 1 (Continued)

Reference	Brain region	*N*	Statistical methodology	Cr-referenced GABA findings	Water-referenced GABA findings
Harada et al, 2011[11]	Lentiform nuclei Left frontal lobe ACC	8	ICC	Overall ICC = 0.72	—
Geramita et al, 2011[14]	ACC Right frontal white matter (rFWM)	10	ICC CV	ACC: ICC = 0.50; CV = 6.5% rFWM: ICC = −0.35; CV = 8.2%	ACC: ICC = 0.79; CV = 5.3% rFWM: ICC = 0.48; CV = 8.7%
O'Gorman et al, 2011[26]	dlPFC	14	CV	—	CV = 7%
Bogner et al, 2010[12]	OCC	11	CV	Processed with integration: CV = 17.0% Processed with fitting: CV = 13.3%	Processed with integration: CV = 17.8% Processed with fitting: CV = 15.0%
Evans et al, 2010[19]	OCC Precentral gyrus	8	CV	—	OCC: CV = 6.5% Precentral gyrus: CV = 8.8%

Note: For the purposes of this table, "water-referenced GABA findings" include studies utilizing an internal water reference and/or water scaling factor.

Abbreviations: ACC, anterior cingulate cortex; BA, Bland-Altman; CV, coefficient of variation; dlPFC, dorsolateral prefrontal cortex; GABA, gamma-aminobutyric acid; GABA+, GABA collected with macromolecular contaminants; GABA', GABA collected with macromolecular suppression; ICC, intraclass correlation coefficient; MEGA-PRESS, Meshcher-Garwood point resolved spectroscopy; MRS, magnetic resonance spectroscopy; OCC, occipital cortex; PCC, posterior cingulate cortex; PFC, prefrontal cortex; ROIs, regions of interest.

[a]These papers were both based on one large dataset from a study across 24–25 different research sites.

were recruited from the greater Boston community and were between the ages of 18 and 25 (M = 21.2 years, SD = 2.4 years), split evenly by sex (51% female). The self-identified racial makeup of the sample was 51% White, 29% Asian, 9% Black, and 9% biracial, with 3% (one participant) declining to answer. Eleven percent identified as Hispanic or Latinx.

All participants were right-handed with no significant medical history or use of psychotropic medications. Given the well-established impact of alcohol use on GABA concentrations, participants with more than five alcohol-related blackouts or an alcohol use disorder were excluded.[35-38] An initial screening visit was conducted wherein written informed consent was obtained in compliance with the requirements of the Partners Human Research Committee. Eligible participants completed a 2-hour MR scan within a month of the screening session. As the menstrual cycle has been shown to impact GABA concentrations in the ACC, all women were scanned during their follicular phase.[39]

2.2 | MRI acquisition

Structural and functional images as well as MRS were acquired via a Siemens 3T Prisma scanner (Siemens Healthcare, Erlangen, Germany) operating at 123 MHz for proton imaging and spectroscopy, using a 64-channel, phased-array head coil for reception and a body coil for transmission. A set of T_1-weighted high-resolution 3D structural images were collected with a multiecho magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence (pulse repetition time [TR] = 2530 ms; echo time [TE] = 1.69, 3.55, 5.41, 7.27 ms; slice thickness = 1 mm; total number of slices = 176; flip angle = 7.0°; field of view = 256 mm; voxel dimensions = 1 × 1 × 1 mm).

The T_1-weighted structural images were used to place a voxel in the rACC (17.5 mL; 35 × 20 × 25 mm^3) and left dlPFC (18.75 mL; 25 × 30 × 25 mm^3) for MRS data collection (Figure 1). Data were collected from both voxels for all participants, with 26 participants having test-retest data collected for one voxel and nine participants for both (depending on scan time constraints). Test-retest data (in the same scanning session) were acquired for 24 participants in the rACC and 20 in the dlPFC. MRS data were collected immediately after a high-resolution localizer (namely the 2D high-resolution images repartitioned from the MPRAGE images) and before any other scans, such that no frequency drift was seen because of gradient heating.

Proton GABA+ (macromolecular-contaminated) measurement employed a Meshcher-Garwood point resolved spectroscopy (MEGA-PRESS) sequence obtained from the University of Minnesota with the acquisition frequency sitting at 3.0 ppm and frequency-selective editing pulses, each with a duration of 17 ms,[40] alternatively at 1.9 ppm (on) and 7.5 ppm (off) interleaved with the averages.[7,40-42] MEGA-PRESS is an established MRS acquisition protocol for GABA detection that has shown superior GABA test-retest reliability compared with other sequences.[8,43] Shimming of

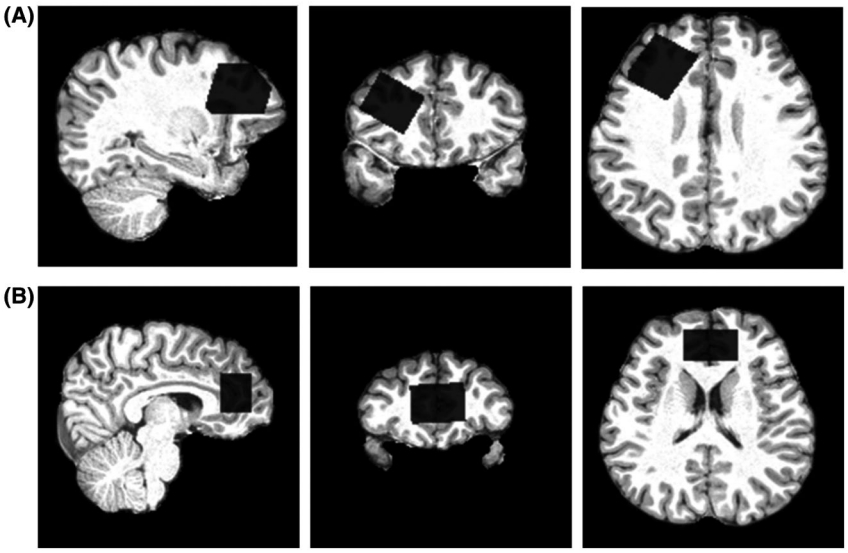

FIGURE 1 Images illustrating the voxel placement for the (A) left dorsolateral prefrontal cortex and (B) rostral anterior cingulate cortex. Voxel placement is presented in sagittal, coronal, and axial views on a single subject for each region

the magnetic field within the prescribed voxel was performed using a vendor-provided 3D shimming routine designated for the human brain region followed by manual adjustment as needed (performed by the same MRS physicist [CSZ] for all participants). The full width at half maximum (FWHM) was measured from both the Siemens console (rACC: M = 17.47 Hz, SD = 1.92 Hz; dlPFC: M = 15.95 Hz, SD = 2.14 Hz) and the unsuppressed water peak (rACC: M = 7.82 Hz, SD = 1.21 Hz; dlPFC: M = 7.69 Hz, SD = 1.27 Hz). A variable pulse power and optimized relaxation (VAPOR) delays module was utilized to achieve water suppression.[44] Following shimming, carrier frequency was adjusted, flip angles and water suppression were optimized, and the MEGA-PRESS spectra were collected at TE = 68 ms, TR = 3000 ms, spectral bandwidth = 1.2 kHz, 2048 data points, readout duration = 1706 ms, total number of signal averages = 192, total scan duration = 10 minutes with applied directions of the slice-selecting gradients identical across subjects (sagittal, R→L; coronal, A→P; transversal, F→H). The test-retest data acquisitions took place back-to-back, with no delay or reacquisition of anatomical images between data collections. Therefore, correction for the voxel tissue composition was not necessary as the voxel locations were identical between each participant's test-retest scans. After MEGA-PRESS data collection, unsuppressed water signal positioned at the location of the GABA 3 ppm resonance was collected with an offset frequency of −1.7 ppm for eddy current correction and quantification purposes, after accounting for the error of chemical shift displacement (≤2%).

2.3 | GABA+ quantification

The MRS data were exported in .IMA format and processed using FID-A.[45] To quantify neural GABA+ concentrations, the 96 edit-on and 96 edit-off FIDs were corrected for possible frequency and phase drift, Gaussian filtered (2 Hz) and Fourier transformed prior to grouping on and off spectra and taking the corresponding edit-on and edit-off spectra differences. The grouped edit-off and difference spectra, in conjunction with corresponding unsuppressed water signals, were imported into the linear combination model (LCModel, version 6.3-1N) to fit the following metabolites: total creatine (tCr), total choline, glutamate, glutamate + glutamine, myoinositol, and N-acetylaspartate (NAA) from the edit-off spectrum, and GABA+ from the difference spectrum (Figure 2). The control parameter sptype, used to indicate special types of spectra, was set to "mega-press-3" for GABA+ quantification. The baseline was stiff and flat following the LCModel default settings. The difference spectra were fitted with an in-house simulated basis set of metabolites using GAMMA.[47] No basis was utilized for the coedited 3 ppm macromolecule signal, as the overlap with GABA could result in unreliability in fitting.

GABA+ concentrations are reported as GABA+/tCr (a ratio of GABA+ to total creatine) and GABA+/water (a ratio of GABA+ to water multiplied by a scaling factor, reported in mM). LCModel fitting of the MRS data was assessed for quality based on Cramer-Rao lower-bound values of <15% and signal-to-noise ratios of >20, with one participant excluded. Additionally, spectra were visually assessed prior to analyses by MR physicists (XC and FD) for severe baseline distortion, excluding two additional participants (final rACC $n = 21$ and dlPFC $n = 20$).

2.4 | Statistical analyses

Statistical analyses were performed using SPSS statistics software (version 24.0; SPSS, Inc., Chicago, IL). To assess closeness of agreement, within-subject CVs were calculated

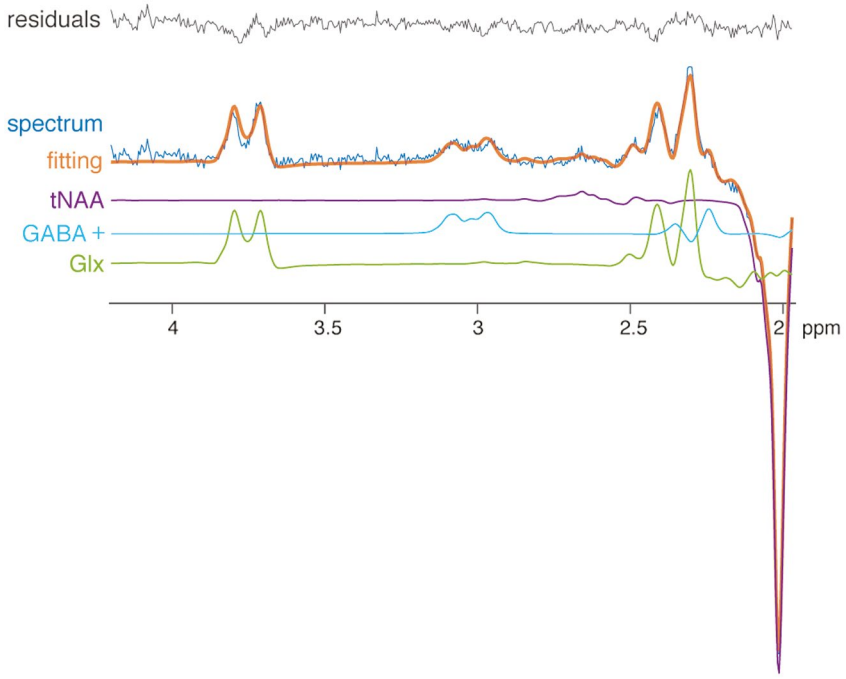

FIGURE 2 GABA+-edited (difference) spectrum showing metabolite fitting lines as estimated with LCModel, depicting the GABA+-edited spectrum (dark blue), fitting line (orange), total N-acetylaspartate (tNAA; purple), GABA+ (light blue), glutamate+glutamine (Glx; green), and residuals (gray). GABA, gamma-aminobutyric acid; GABA+, GABA collected with macromolecular contaminants; LCModel, linear combination model

by first computing the SD of the paired measurements divided by the mean for each subject, and then averaging across participants for each voxel:

$$\mathrm{CV} = \frac{SD}{\bar{x}} \times 100 = \frac{\sqrt{\sum_{i=1}^{i=2} (x_i - \bar{x})^2}}{\bar{x}} \times 100 \quad (1)$$

where x_i is the value of the ith scan for each participant and SD is the sample SD for each participant. CVs are the most commonly reported metric in the MRS test-retest literature,[8,48] and provide an assessment of within-subject measurement agreement that is independent of the range of values in the sample.

Next, Bland-Altman (BA) plots were created to visually examine measurement repeatability.[49-51] BA plots provide a useful complement to agreement metrics such as CVs, allowing for easy identification of outliers and systematic trends in measurement error. Following Bland and Altman,[49,50] the difference between the paired measurements was plotted against the mean for each participant (BlandAltmanLeh package in R, version 3.5.3).[52] One-sample t tests were used to check the assumption that this mean difference was not significantly different from zero.[49,50]

To calculate limits of agreement (between which 95% of the data points are expected to lie) Bland and Altman[49,50] recommend using repeatability coefficients for test-retest studies. These limits assume an interscan mean difference of zero, which is most appropriate for studies using the same measurement technique on the same subjects, provided it is reasonable to assume no systematic mean difference in observations over time. Thus, repeatability coefficients were calculated as the t critical value (for the upper and lower 2.5% tails) for each voxel multiplied by the SD of the interscan differences, with an assumed mean difference set at zero[49,50] as follows:

$$\text{Repeatability coefficient} = \pm t_{0.975} \sqrt{\frac{\sum_{j=1}^{j=n} \left(x_{1j} - x_{2j}\right)^2}{n-1}} \quad (2)$$

where j indexes the individual participants, and n = total number of participants for each voxel. The repeatability limits of agreement and a zero line of no difference were added to the plots to aid in the assessment of the magnitude of measurement error and identify outliers.

To investigate reliability, single-rating, absolute-agreement, two-way mixed-effects ICCs and their 95% CIs were calculated.[53] ICCs assess how reliably an instrument distinguishes between subjects[54] and are calculated as the between-subject variance divided by the total variance. They are widely used by different fields of study to measure reliability.[8,53-55]

3 | RESULTS

Within the dlPFC, the mean CV for GABA+/tCr and GABA+/water were 4.6% and 4.0%, respectively. GABA+/tCr and GABA+/water produced ICCs of 0.77 and 0.87, respectively (Table 2). Both the GABA+/tCr and GABA+/water BA plots showed 95% of participants within the repeatability coefficients (one outlier each; Figure 3), as is expected

TABLE 2 GABA+/tCr and GABA+/water test-retest results

		CV		ICC	
Voxel & ratio	*n*	Mean (%)	Range (%)	Value	95% CI
dlPFC GABA+/tCr	20	4.60	0.03–11.68	0.77	0.51–0.90
dlPFC GABA+/water	20	3.97	0.24–11.02	0.87	0.70–0.95
rACC GABA+/tCr	21	8.05	1.27–21.99	0.40	0.01–0.70
rACC GABA+/water	21	7.49	0.15–22.68	0.44	0.05–0.72

Abbreviations: CV, coefficient of variation; dlPFC, dorsolateral prefrontal cortex; GABA, gamma-aminobutyric acid; GABA+, GABA collected with macromolecular contaminants; ICC, intraclass correlation coefficient; rACC, rostral anterior cingulate cortex; tCr, total creatine.

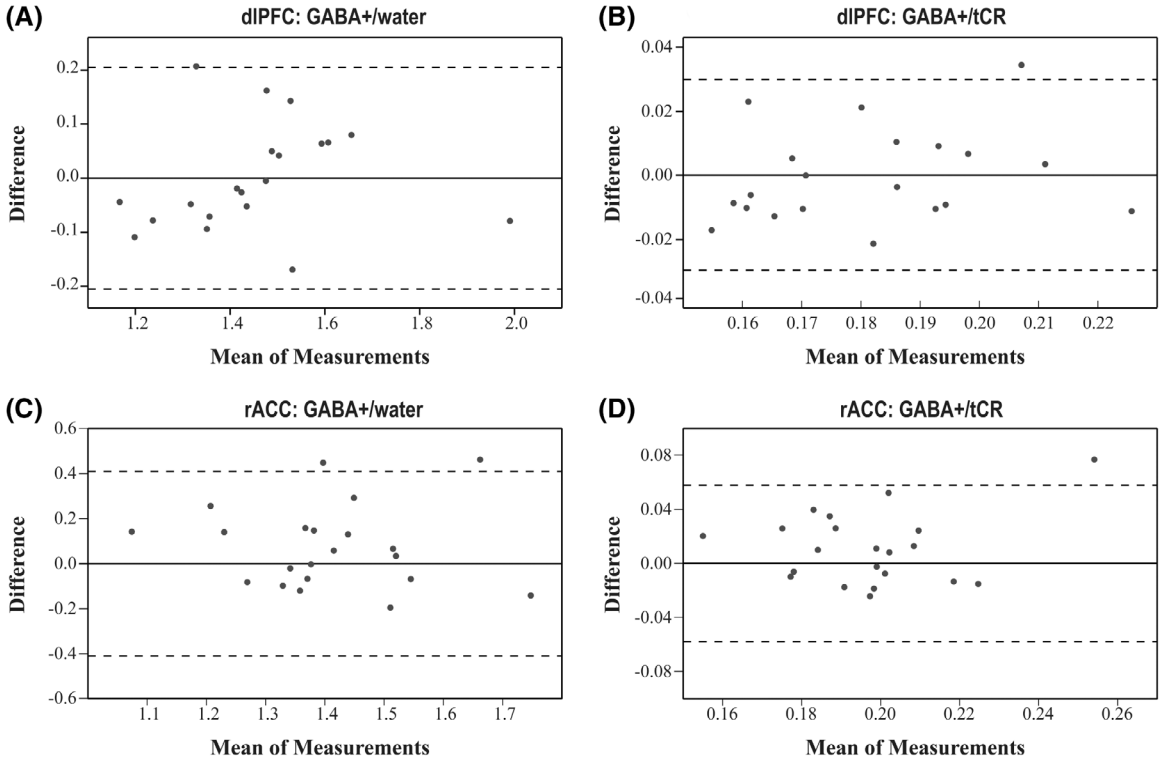

FIGURE 3 Bland-Altman plots for (A) dlPFC: GABA+/water, (B) dlPFC: GABA+/tCR, (C) rACC: GABA+/water, and (D) rACC: GABA+/tCR. The plots depict the mean of scan 1 and scan 2 on the *x* axis and the interscan difference on the *y* axis. The solid line represents the zero line of no difference. The dotted lines depict limits of agreement, calculated with repeatability coefficients (the critical *t* value multiplied by the SD of the interscan differences). dlPFC, dorsolateral prefrontal cortex; GABA, gamma-aminobutyric acid; GABA+, GABA collected with macromolecular contaminants; rACC, rostral anterior cingulate cortex; tCr, total creatine

with unbiased, normally distributed measurement error.[49] For both GABA+/tCr and GABA+/water, the mean interscan differences were not significantly different from zero (M for GABA+/tCr: −0.0004; 95% CI, −0.0072 to 0.0063, P = .89; M for GABA+/water: 0.001; 95% CI, −0.045 to 0.047, P = .96).

Within the rACC, the mean CV for GABA+/tCr and GABA+/water were 8.0% and 7.5%, respectively. The rACC showed lower ICC values, with GABA+/tCr and GABA+/water ICCs of 0.40 and 0.44, respectively. Within the BA plots, 95% of measurements fell between the repeatability coefficients for GABA+/tCr (one outlier), while the GABA+/water plot showed 91% of measurements between the repeatability coefficients (two outliers). The mean interscan differences were again not significantly different from zero (M for GABA+/tCr: 0.009; 95% CI, −0.002 to 0.022], P = .10; M for GABA+/water: 0.07; 95% CI, −0.01 to 0.16, P = .08).

4 | DISCUSSION

The present study found greater GABA+ MRS test-retest repeatability and reliability within the dlPFC as compared with the rACC. These strong findings in the dlPFC replicated results

of past studies in a larger cohort.[18,22,25,26] Although reliability thresholds for ICCs are not well-established within the broader literature, dlPFC ICCs were well above the 0.75 threshold characterized as "excellent" by both Baeshen et al[8] and Cicchetti et al[56] and "good" by Koo and Li,[55] suggesting that MEGA-PRESS can be used to reliably distinguish between the dlPFC GABA+ levels of healthy subjects. Measurement agreement in the dlPFC was also strong. Thresholds for interpreting CVs are rarely discussed in the MRS literature, but the current averaged CVs were far below those previously used (eg, 20% used by Baeshen et al[8]). Additionally, BA plots revealed few outliers with no systematic trends in measurement error, further indicating strong agreement.

Overall, the rACC showed lower agreement and reliability than the dlPFC, possibly related to the nearby sinus cavity affecting shim quality. Indeed, the mean rACC FWHM calculated from both the Siemens console and the final data (peak of the unsuppressed water scan) was higher than in the dlPFC, though this difference was not significant when calculated from the final data ($P > .7$). Measures of repeatability (ie, averaged CVs and BA plots) yielded stronger results than measures of reliability (ie, ICCs) in the rACC, highlighting the importance of reporting both metrics. Averaged CVs showed moderate agreement, and BA plots revealed few outliers and no bias in measurement error. Meanwhile, rACC ICCs were in the low-moderate range.[8,55,56]

To further contextualize the principal measures of agreement and reliability, it is useful to examine the mathematical structure of the CV and ICC. The CV depends exclusively on the within-subject variance, whereas the ICC is dependent on both the between-subject and within-subject variances. Thus, as shown by Bland and Altman,[57] for a given level of measurement repeatability (eg, CV), there can be marked variability in the ICC depending on the range of values within the chosen sample. In the present study of healthy young adults, the somewhat truncated range of observed rACC GABA+ concentrations may, therefore, partially explain the poor ICCs. Furthermore, it should be considered that the dependence of the CV and ICC on the within-subject variance is not symmetric. Specifically, the ICC decreases as the within-subject variance increases, whereas the CV increases in proportion to the square root of the within-subject variance. Thus, increases in CV that reflect a modest increase on the square root scale can have a much larger effect on decreasing the ICC. In the context of the present study, the impact of greater within-subject variance shown in the rACC would, therefore, be magnified in the ICC calculations as compared with those of the CV.

The results of this study should be considered with respect to its limitations, including the use of only two consecutive scans to assess test-retest repeatability and the inclusion of only healthy participants. Keeping the participant in situ between scans obviated the need to reposition the voxel for the second scan, thereby reducing measurement error and likely increasing agreement. Thus, between-group comparisons of GABA concentrations may be subject to greater noise related to voxel placement than shown by these results. In addition, the study sample included only young adults without a history of psychopathology, who may show a narrower range of GABA levels compared with the broader population. As discussed above, this limited variability likely lowered the estimate of reliability based on the ICC.

Future research with a larger sample of psychiatric cases and healthy controls is needed to more thoroughly assess the test-retest reliability of rACC GABA across different population groups. Investigation is also needed to improve techniques for macromolecular suppression with MEGA-PRESS. GABA+ concentrations with MEGA-PRESS in the present study and others are likely overestimated because of macromolecular contamination, and current techniques for suppression of this contamination pose challenges.[7,8] Macromolecules may also vary with factors such as age or brain region, potentially increasing the difficulty of interpreting GABA findings.[58-60] Thus, improving techniques for macromolecular suppression remains a critical area for future work.

Despite these limitations, the present study provides evidence of excellent GABA+ test-retest repeatability and reliability in the dlPFC, and moderate repeatability in the rACC in a larger healthy sample. These results provide important context for future studies of clinical populations and further methodological work.

ACKNOWLEDGMENTS

Jill M. Goldstein and Diego A. Pizzagalli provided equal senior contributions to this work. This project was supported by R01MH108602 (to DAP and JMG) from the National Institute of Mental Health. DAP was partially supported by R37 MH068376. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funding organization had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The MEGA-PRESS sequence was developed by Edward J. Auerbach and Małgorzata Marjańska and provided by the University of Minnesota under a C2P agreement. We would also like to thank Garrett Fitzmaurice, ScD, for his consultation regarding the statistical methods used in this work, and Madeline (Lynn) Alexander, PhD, Laurie A. Scott, AM, and Harlyn Aizley, EdM, for clinical interviews to establish study eligibility.

DISCLOSURE STATEMENT

Over the past 3 years, Dr. Pizzagalli has received consulting fees from Akili Interactive Labs, BlackThorn Therapeutics, Boehringer Ingelheim, Compass Pathway, Otsuka Pharmaceuticals, and Takeda Pharmaceuticals; one

honorarium from Alkermes, and research funding from the NIMH, Dana Foundation, Brain and Behavior Research Foundation, and Millennium Pharmaceuticals. In addition, he has received stock options from BlackThorn Therapeutics. Dr. Hudson has received grant support from Boehringer Ingelheim and Sunovion, and has received consulting fees from Idorsia, Shire, and Sunovion. Dr. Goldstein is on the scientific advisory board and has an equity interest in Cala Health, a neuromodulation device company. There are no conflicts of interest with the work conducted in this study. No funding from these entities was used to support the current work, and all views expressed are solely those of the authors. The other authors have no financial disclosures.

DATA AVAILABILITY STATEMENT

The data presented here are available at the NIMH Data Archive (https://nda.nih.gov/edit_collection.html?id=2485).

ORCID

Jessica M. Duda https://orcid.org/0000-0001-5455-1607
Amelia D. Moser https://orcid.org/0000-0002-9889-6102
Maria Ironside https://orcid.org/0000-0003-2781-1711
David J. Crowley https://orcid.org/0000-0001-9770-4482
Laura M. Holsen https://orcid.org/0000-0003-1292-7789
Madhusmita Misra https://orcid.org/0000-0002-9405-2981
James I. Hudson https://orcid.org/0000-0003-3999-8931
Jill M. Goldstein https://orcid.org/0000-0003-1689-7791
Diego A. Pizzagalli https://orcid.org/0000-0002-7772-1143

TWITTER

Diego A. Pizzagalli @DiegoPizzagalli

REFERENCES

1. Egerton A, Modinos G, Ferrera D, McGuire P. Neuroimaging studies of GABA in schizophrenia: A systematic review with meta-analysis. *Transl Psychiatry*. 2017;7:e1147.
2. Godfrey KEM, Gardner AC, Kwon S, Chea W, Muthukumaraswamy SD. Differences in excitatory and inhibitory neurotransmitter levels between depressed patients and healthy controls: A systematic review and meta-analysis. *J Psychiatr Res*. 2018;105:33-44.
3. Luscher B, Shen Q, Sahir N. The GABAergic deficit hypothesis of major depressive disorder. *Mol Psychiatry*. 2011;16:383-406.
4. Schür RR, Draisma LW, Wijnen JP, et al. Brain GABA levels across psychiatric disorders: A systematic literature review and meta-analysis of 1H-MRS studies. *Hum Brain Mapp*. 2016;37:3337-3352.
5. Rajkowska G, O'Dwyer G, Teleki Z, Stockmeier CA, Miguel-Hidalgo JJ. GABAergic neurons immunoreactive for calcium binding proteins are reduced in the prefrontal cortex in major depression. *Neuropsychopharmacology*. 2007;32:471-482.
6. Nakazawa K, Zsiros V, Jiang Z, et al. GABAergic interneuron origin of schizophrenia pathophysiology. *Neuropharmacology*. 2012;62:1574-1583.
7. Mullins PG, McGonigle DJ, O'Gorman RL, et al. Current practice in the use of MEGA-PRESS spectroscopy for the detection of GABA. *Neuroimage*. 2014;86:43-52.
8. Baeshen A, Wyss PO, Henning A, et al. Test–retest reliability of the brain metabolites GABA and Glx with JPRESS, PRESS, and MEGA-PRESS MRS sequences in vivo at 3T. *J Magn Reson Imaging*. 2020;51:1181-1191.
9. 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.
10. Mikkelsen M, Singh KD, Sumner P, Evans CJ. Comparison of the repeatability of GABA-edited magnetic resonance spectroscopy with and without macromolecule suppression. *Magn Reson Med*. 2016;75:946-953.
11. Harada M, Kubo H, Nose A, Nishitani H, Matsuda T. Measurement of variation in the human cerebral GABA level by in vivo MEGA-editing proton MR spectroscopy using a clinical 3 T instrument and its dependence on brain region and the female menstrual cycle. *Hum Brain Mapp*. 2011;32:828-833.
12. 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.
13. Near J, Ho Y-CL, Sandberg K, Kumaragamage C, Blicher JU. Long-term reproducibility of GABA magnetic resonance spectroscopy. *Neuroimage*. 2014;99:191-196.
14. 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.
15. Napolitano A, Kockenberger W, Auer DP. Reliable gamma aminobutyric acid measurement using optimized PRESS at 3 T. *Magn Reson Med*. 2013;69:1528-1533.
16. Prinsen H, de Graaf RA, Mason GF, Pelletier D, Juchem C. Reproducibility measurement of glutathione, GABA, and glutamate: Towards in vivo neurochemical profiling of multiple sclerosis with MR spectroscopy at 7T. *J Magn Reson Imaging*. 2017;45:187-198.
17. Prisciandaro JJ, Mikkelsen M, Saleh MG, Edden RAE. An evaluation of the reproducibility of 1H-MRS GABA and GSH levels acquired in healthy volunteers with J-difference editing sequences at varying echo times. *Magn Reson Imaging*. 2020;65:109-113.
18. Wijtenburg SA, Rowland LM, Edden RAE, Barker PB. Reproducibility of brain spectroscopy at 7T using conventional localization and spectral editing techniques. *J Magn Reson Imaging*. 2013;38:460-467.
19. Evans CJ, McGonigle DJ, Edden RAE. Diurnal stability of γ-aminobutyric acid concentration in visual and sensorimotor cortex. *J Magn Reson Imaging*. 2010;31:204-209.
20. Near J, Andersson J, Maron E, et al. Unedited in vivo detection and quantification of γ-aminobutyric acid in the occipital cortex using short-TE MRS at 3T. *NMR Biomed*. 2013;26:1353-1362.
21. Greenhouse I, Noah S, Maddock RJ, Ivry RB. Individual differences in GABA content are reliable but are not uniform across the human cortex. *Neuroimage*. 2016;139:1-7.
22. Yasen AL, Smith J, Christie AD. Reliability of glutamate and GABA quantification using proton magnetic resonance spectroscopy. *Neurosci Lett*. 2017;643:121-124.
23. Gaetz W, Bloy L, Wang DJ, et al. GABA estimation in the brains of children on the autism spectrum: Measurement precision and regional cortical variation. *Neuroimage*. 2014;86:1-9.

24. Ferland MC, Therrien-Blanchet JM, Lefebvre G, Klees-Themens G, Proulx S, Théoret H. Longitudinal assessment of 1H-MRS (GABA and Glx) and TMS measures of cortical inhibition and facilitation in the sensorimotor cortex. *Exp Brain Res*. 2019;237:3461-3474.
25. Shungu DC, Mao X, Gonzales R, et al. Brain γ-aminobutyric acid (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;29:932-942.
26. 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.
27. Mikkelsen M, Rimbault DL, Barker PB, et al. Big GABA II: Water-referenced edited MR spectroscopy at 25 research sites. *NeuroImage*. 2019;191:537-548.
28. Mikkelsen M, Barker PB, Bhattacharyya PK, et al. Big GABA: Edited MR spectroscopy at 24 research sites. *Neuroimage*. 2017;159:32-45.
29. Long Z, Dyke JP, Ma R, Huang CC, Louis ED, Dydak U. Reproducibility and effect of tissue composition on cerebellar GABA MRS in an elderly population. *NMR Biomed*. 2015;28:1315-1323.
30. Juchem C, de Graaf RA. B0 magnetic field homogeneity and shimming for in vivo magnetic resonance spectroscopy. *Anal Biochem*. 2017;529:17-29.
31. Koch KM, Rothman DL, de Graaf RA. Optimization of static magnetic field homogeneity in the human and animal brain in vivo. *Prog Nucl Magn Reson Spectrosc*. 2009;54:69-96.
32. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: Analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol*. 2008;31:466-475.
33. Kottner J, Audige L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011;64:96-106.
34. First MB, Williams JBW, Karg RS, Spitzer RL. *Structured Clinical Interview for DSM-5, Research Version*. Arlington, VA: American Psychiatric Association; 2015.
35. Abé C, Mon A, Durazzo TC, Pennington DL, Schmidt TP, Meyerhoff DJ. Polysubstance and alcohol dependence: Unique abnormalities of magnetic resonance-derived brain metabolite levels. *Drug Alcohol Depend*. 2013;130:30-37.
36. Gomez R, Behar KL, Watzl J, et al. Intravenous ethanol infusion decreases human cortical γ-aminobutyric acid and N-acetylaspartate as measured with proton magnetic resonance spectroscopy at 4 Tesla. *Biol Psychiatry*. 2012;71:239-246.
37. Prisciandaro JJ, Schacht JP, Prescot AP, Renshaw PF, Brown TR, Anton RF. Brain glutamate, GABA, and glutamine levels and associations with recent drinking in treatment-naïve individuals with alcohol use disorder versus light drinkers. *Alcohol Clin Exp Res*. 2019;43:221-226.
38. Silveri MM, Cohen-Gilbert J, Crowley DJ, Rosso IM, Jensen JE, Sneider JT. Altered anterior cingulate neurochemistry in emerging adult binge drinkers with a history of alcohol-induced blackouts. *Alcohol Clin Exp Res*. 2014;38:969-979.
39. Silveri MM, Sneider JT, Crowley DJ, et al. Frontal lobe γ-aminobutyric acid levels during adolescence: Associations with impulsivity and response inhibition. *Biol Psychiatry*. 2013;74:296-304.
40. Marjańska M, Lehéricy S, Valabrègue R, et al. Brain dynamic neurochemical changes in dystonic patients: A magnetic resonance spectroscopy study. *Mov Disord*. 2013;28:201-209.
41. Mescher M, Merkle H, Kirsch J, Garwood M, Gruetter R. Simultaneous in vivo spectral editing and water suppression. *NMR Biomed*. 1998;11:266-272.
42. Tremblay S, Beaulé V, Proulx S, et al. The use of magnetic resonance spectroscopy as a tool for the measurement of bi-hemispheric transcranial electric stimulation effects on primary motor cortex metabolism. *J Vis Exp*. 2014.
43. 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. *J Magn Reson*. 2011;208:210-218.
44. Tkáč I, Starčuk 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.
45. Simpson R, Devenyi GA, Jezzard P, Hennessy TJ, Near J. Advanced processing and simulation of MRS data using the FID appliance (FID-A)—An open source, MATLAB-based toolkit. *Magn Reson Med*. 2017;77:23-33.
46. Provencher SW. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magn Reson Med*. 1993;30:672-679.
47. Smith SA, Levante TO, Meier BH, Ernst RR. Computer simulations in magnetic resonance. An object-oriented programming approach. *J Magn Reson Ser A*. 1994;106:75-105.
48. Gasparovic C, Bedrick EJ, Mayer AR, et al. Test-retest reliability and reproducibility of short-echo-time spectroscopic imaging of human brain at 3T. *Magn Reson Med*. 2011;66:324-332.
49. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327:307-310.
50. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135-160.
51. van Stralen KJ, Jager KJ, Zoccali C, Dekker FW. Agreement between methods. *Kidney Int*. 2008;74:1116-1120.
52. R Core Team. *R: A Language and Environment for Statistical Computing*. Published online. 2019. https://www.r-project.org/
53. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 1979;86:420-428.
54. Shoukri MM, Colak D, Kaya N, Donner A. Comparison of two dependent within subject coefficients of variation to evaluate the reproducibility of measurement devices. *BMC Med Res Methodol*. 2008;8.
55. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155-163.
56. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6:284-290.
57. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med*. 1990;20:337-340.
58. Aufhaus E, Weber-Fahr W, Sack M, et al. Absence of changes in GABA concentrations with age and gender in the human anterior cingulate cortex: A MEGA-PRESS study with symmetric editing pulse frequencies for macromolecule suppression. *Magn Reson Med*. 2013;69:317-320.

59. Hofmann L, Slotboom J, Boesch C, Kreis R. Characterization of the macromolecule baseline in localized 1H-MR spectra of human brain. *Magn Reson Med*. 2001;46:855-863.
60. Mader I, Seeger U, Karitzky J, Erb M, Schick F, Klose U. Proton magnetic resonance spectroscopy with metabolite nulling reveals regional differences of macromolecules in normal human brain. *J Magn Reson Imaging*. 2002;16:538-546.

How to cite this article: Duda JM, Moser AD, Zuo CS, et al. Repeatability and reliability of GABA measurements with magnetic resonance spectroscopy in healthy young adults. *Magn Reson Med*. 2021;85:2359–2369. https://doi.org/10.1002/mrm.28587