ORIGINAL ARTICLE

Seeking Ground Truth for GABA Quantification by Edited Magnetic Resonance Spectroscopy: Comparative Analysis of TARQUIN, LCModel, JMRUI and GANNET

C. Jenkins^{1,2,3*} | M. Chandler^{4*} | F. C. Langbein^{4*}

S. M. Shermer^{1,2}*

¹College of Science (Physics), Swansea University, Swansea, SA2 8PP, United Kingdom

²Centre for Nanohealth & Clinical Imaging Unit, Institute for Life Science, Swansea University, Swansea, SA2 8PP, United Kingdom

³Cardiff University Brain Research Imaging Centre (CUBRIC), Maindy Rd, Cardiff, CF24 4HQ, United Kingdom

⁴School of Computer Science and Informatics, Cardiff University, Cardiff, CF24 3AA, United Kingdom

Correspondence S. M. Shermer, PhD Email: lw1660@gmail.com

Funding information

Royal Society Leverhulme Trust Senior Fellowship (SMS); ABMU Health Board, College of Science, Welcome trust strategic award (CJ); EPSRC studentship (MC) **Purpose:** Many tools exist for quantifying magnetic resonance spectroscopy (MRS) data. Literature comparing them is sparse but indicates potential methodological bias. We benchmark MRS analysis tools to elucidate this.

Methods: Four series of phantom experiments, including both solutions and tissue-mimicking gels, with constant concentrations of NAA, Creatine, Glutamine and Glutamate, and iteratively increased concentrations of GABA are performed. MEGA-PRESS spectra are acquired and quantified with several state-of-the-art MRS analysis tools (LCModel, TARQUIN, JMRUI, GANNET) and in-house code (LWFIT). GABA-to-NAA ratios for reported metabolite amplitudes are compared to the ground truth of known concentration ratios. Overall estimation accuracy is assessed by linear fits of reported vs. actual ratios and coefficients of determination. Simulations further elucidate the experimental results.

Results: Significant differences in reported GABA-to-NAA amplitude ratios are observed. TARQUIN consistently overestimates, while most tools underestimate ratios to vary-

* Equally contributing authors.

ing degrees compared to the ground truth. Underestimation due to reduced editing efficiency is predicted by simulations. LCModel performs comparatively well for wellresolved solution spectra but struggles for intentionally miscalibrated spectra and gel spectra mimicking *in-vivo* conditions. GANNET shows better consistency and robustness to calibration errors but greater underestimation related to how the 3 ppm GABA peak is fitted. Surprisingly, simple peak integration with minimal pre-processing yields the most consistent and accurate results compared to the ground truth.

Conclusions: A methodological dependence is observed not only in the quantification results for individual spectra, but in GABA-to-NAA gradients across experimental series, systematic offsets and coefficients of determination.

KEYWORDS MRS, GABA, MEGA-PRESS, LCModel, TARQUIN, JMRUI, GANNET

1 | INTRODUCTION

Gamma-aminobutyric acid (GABA) is the primary inhibitory neurotransmitter in the central nervous system, maintaining the excitation-inhibition balance. Its prominent role in neurotransmission and metabolism has led to extensive study and a plethora of applications for its detection. Disruption of GABAergic processes has been observed in Schizophrenic patients [22, 47, 53] and GABA receptor dysfunction has associations with epilepsy [18, 31]. GABA processes are prominent in type I diabetes [49] and autism spectrum disorders [6]. GABA is also the subject of intense study by the psychological community, with GABA levels influencing impulsivity [5, 9], drug addiction [7],

anxiety disorders [41] and depression [43]. However, despite its wide range of applications, there are many challenges associated with its detection and quantification *in-vivo*. The most common tool for non-invasive detection of GABA *in-vivo* is magnetic resonance spectroscopy (MRS). However, as GABA exists in the brain in mmol concentrations and relatively stronger signals from other metabolites present overlap with, and obscure, the GABA signal, GABA becomes generally observable at clinically relevant magnetic field strengths only through the use of edited MRS [11, 24, 25, 27] or other specialised techniques, such as 2D *J*-spectroscopy [15].

The most commonly used MRS pulse sequence to detect GABA at fields strength up to 3 T is the MEGA-PRESS sequence [23, 25, 37, 54], which uses the PRESS [4] localisation scheme with the addition of a pair of frequency selective editing pulses, placed symmetrically about the second refocusing pulse. Two acquisitions — 'edit on' and 'edit off' — are made, which differ only in the frequency of the editing pulses. The frequency is chosen to take advantage of the *J*-evolution of the target metabolites, such that upon subtraction of the two acquisitions, only the resonances in the edit bandwidth, and those coupled to them, remain. The widespread use of MRS, particularly edited MRS,

has seen the development of a range of software packages aimed at processing and quantification of the spectra [8, 28, 35, 36, 40, 52, 56]. Quantification of MRS data is usually performed after pre-processing, which entails some combination of apodization, frequency and phase calibration, macromolecular baseline subtraction and residual water signal removal. Each step may be realised by a diverse array of methods, providing many potential avenues of variation for the quantified signal. This effect is compounded by further differences in the quantification procedure. The most common approach to quantification is the generation of a set of basis spectra — via phantom scans or simulation — followed by the application of a fitting algorithm to decompose the MRS signal into weighted combinations of the basis functions. However, the basis functions and fit algorithms vary by analysis tool, which provides potential for methodological bias.

While it has been argued that different methods, suitably applied, exhibit similar variation and clinical observations [42, 44], it is acknowledged in the literature that there is procedural dependency in MRS analysis [14, 19, 34], which provides a statistically significant variation in reported metabolite amplitudes or relative concentrations [3, 21]. The goal of this study is to assess the variation of quantification results obtained, and to generate benchmark data sets for the purpose of assessing the performance of new and existing tools. Previous work in this area has utilised either *in-vivo* data – for which there is usually no ground truth to evaluate the results – or simulated data sets, which inevitably cannot capture all sources of experimental variation. For this study, a series of MRS data sets based on calibrated phantoms were generated [45]. While the phantoms inevitably fail to capture all influences of the *in-vivo* environment, they do allow the inclusion of many experimental factors, while still enabling a ground truth comparison and some ability to control environmental factors expected to influence results, allowing elucidation of their effect.

Four series of phantom experiments were conducted. In each series, the concentrations of various metabolites – creatine (Cr), N-acetyl-L-aspartic acid (NAA), glutamine (Gln) and glutamate (Glu) – are fixed at approximately *in-vivo* levels, while the concentration of GABA is iteratively increased. MEGA-PRESS spectra are acquired for each concentration step and the GABA-to-NAA ratios are calculated based on metabolite amplitudes reported by different popular MRS analysis software. NAA was chosen as a reference metabolite due to its prominence *in-vivo* and its consistent presence in difference spectra, unlike Cr and water, the signals of which should be eliminated in the difference spectra. The four series of experiments were designed to increase the complexity of the spectra, starting with pH and temperature calibrated solutions containing only NAA, Cr and GABA (E1), adding Glu and Gln to the solutions to increase the complexity of the spectra (E3) and finally progressing to a series of gel phantoms to more closely mimic *in-vivo* spectra (E4)¹. To investigate the effect of calibration errors, a series of experiments with intentionally miscalibrated solutions of NAA, Cr and GABA (E2) were also performed. The GABA-to-NAA ratios derived from the metabolite amplitudes reported by different tools are plotted against the known concentration ratios, and data fitting using linear regression is performed to extract the gradient, offset, and the coefficient of determination, *R*². In an ideal setting we expect the ratio of the metabolite amplitudes to scale linearly (*R*² = 1) with the actual concentration ratio, with a gradient of 1 and zero offset corresponding to perfect quantification of GABA relative to NAA.

2 | MATERIALS AND METHODS

Phantom preparation procedures, experimental setup and scan parameters, and details of the analysis are described.

¹E4 corresponds to E4a in full dataset released.





(a) Solution phantom.

(b) Spherical gel phantom.

FIGURE 1 Images of phantoms. The gel phantoms are approximately 5.8 cm in diameter and the diameter of the flask is approximately 8.2 cm.

TABLE 1 Concentration information in units of mM = mmol/l. All series, except E2, are pH calibrated to 7.0 ± 0.2 .

Series	NAA	Cr	Glu	Gln	GABA
E1	15.00	8.00	0.00	0.00	0.00, 0.52, 1.04, 1.56, 2.07, 2.59, 3.10, 4.12, 6.15, 8.15, 10.12, 11.68
E2	15.00	8.00	0.00	0.00	0.00, 0.50, 1.00, 1.50, 2.00, 2.50, 3.00, 3.99, 4.98, 5.96, 6.95, 7.93, 8.90, 9.88, 11.81
E3	15.00	8.00	12.00	3.00	0.00, 1.00, 2.00, 3.00, 3.99, 4.98, 5.97, 6.95, 7.93, 8.91, 9.88, 10.85, 11.81, 12.77, 13.73
E4	15.00	8.00	12.00	3.00	0.00, 1.00, 2.00, 3.00, 4.00, 6.00, 8.00, 10.00

2.1 | Phantom preparation

The phantom study consists of four series of experiments – E1, E2, E3 and E4 – where the GABA (CAS-20791) concentration of a metabolite phantom is varied over multiple acquisitions using a fixed MRS protocol. The background metabolites – NAA (CAS-997-55-7), Cr (CAS-6020-87-7), Glu (CAS-142-47-2) and Gln (CAS-56-85-9) – are maintained at a fixed concentration for a given series, so that GABA gradients are purely a result of GABA concentration changes. E1, E2 and E3 were solution series, where the GABA concentration of a liquid phantom was increased incrementally. For series E4, several spherical gel phantoms were made with varying GABA concentration. The full concentration information of the experimental series is listed in Table 1 and representative images of the phantoms are shown in Figure 1. A detailed description of the phantom preparation is included in Appendix A.

2.2 | Scan Protocols and Data Acquisition

All MRS scans were conducted at Swansea University's Clinical Imaging Facility using a Siemens 3 T Magnetom Skyra. The scanner room was temperature controlled to $20 \pm 0.6^{\circ}$ C. Signal acquisition was done using the built-in spine coils, specifically the four channel spine coil element 'SP2'. The spine coils provided the platform for the most reproducible set-up and SP2 exhibited the highest SNR of the spine coil elements available. The phantom was aligned with this

element and raised to isocenter, using a cardboard phantom holder, to ensure maximum field homogeneity for the scans. Double-spin-echo field maps were acquired for each phantom to assess the homogeneity and manual shimming and frequency calibration was performed to optimise the spectral width. Further manual calibration was performed to adjust the TX reference voltage and suppression pulses.

A $20 \times 20 \times 20 \text{mm}^3$ voxel at the isocenter of the magnet was selected for single voxel spectroscopy. The MEGA-PRESS spectra were acquired using the Siemens WIP MEGA-PRESS pulse sequence with CHESS water-suppression [30] with $T_R = 2000 \text{ ms}$, $T_E = 68 \text{ ms}$, N = 160 averages and a sampling frequency of 1250 Hz, N = 2048 samples. Editing pulses were applied at 1.9 ppm during the on acquisition and 7.4 ppm during the off acquisition. Spectra with water suppression turned off were also acquired to calculate linewidth and water suppression factors and for water referencing.

For each experiment, the raw data acquired, the single-average coil-combined spectra and combined average edit-on, edit-off and difference spectra produced by the vendor-supplied spectroscopy software are available [45]. Non-water-suppressed spectra for Eddy current correction and internal water referencing were also acquired. For the comparison of the different software packages only the combined-average edit-on, edit-off and difference spectra produced by the vendor-supplied spectroscopy software in Siemens dicom (IMA) format were used, as this format was universally supported by all the software packages evaluated, allowing a more encompassing cross-section of the methods. The common input format also reduced the potential sources of variation to the fitting and quantification procedure (see Appendix B.2 for further information).

2.3 | Analysis methods

Prior to quantification, the 2 ppm NAA and water peaks in each spectrum are fitted by Lorentzians, from which the the position, full-width-half-maximum (FWHM) and maximum signal amplitude of the resonances are obtained. The noise floor was estimated by computation of the standard deviation, σ , of the signal in the region 8–9 ppm, observed to be free of metabolite signal, then the signal-to-noise ratio (SNR) calculated: $\frac{S}{2\sigma}$. Water suppression factors are estimated by comparing the height and area of the water peak in the water unsuppressed and water-suppressed spectra.

The analysis software considered in this study can be broadly separated into two categories: (i) basis set methods (TARQUIN, JMRUI, LCModel), which attempt to fit the data using a set of simulated or experimentally obtained basis spectra and (ii) peak fitting and integration tools (GANNET, LWFIT), which compute the area of peaks associated with various metabolites. The basis set methods generally use a least-squares (LS) or non-linear least-squares (NLLS) approach to fit spectral data. Each analysis tool returns a set of metabolite amplitudes that indicate the relative contributions of various metabolites to the overall spectrum. As the absolute signal values are arbitrary and depend on the scanner design, transmitter calibration, coils used and the phantom itself, we focus on quantifying the relative contributions of different metabolites. To this end, the metabolite amplitudes reported by each tool are plotted vs. the metabolite concentrations and linear fits minimising the least-squares error are performed using the MATLAB curve-fitting toolbox, focusing in particular on the ratio of the GABA and NAA amplitudes vs. the corresponding concentration ratios. Brief descriptions of each tool included in this study are given in Appendix C and readers are directed to the referenced literature for further information.

2.4 | Simulations

To facilitate interpretation of the experimental results, simulations were performed to emulate the experimental setup, modulating GABA concentration while maintaining a fixed concentration of other selected metabolites. MEGA-PRESS

	Δν	$\sigma(v)$	ΔTX	$\sigma(TX)$	ΔG_x	$\sigma(G_x)$	ΔG_y	$\sigma(G_y)$	ΔG_z	$\sigma(G_z)$
E1	22	6.02	13.4	4.41	37	9.25	92	33.13	19	7.03
E2	28	7.42	17.1	4.89	36	10.58	51	16	62	16.52
E3	3	0.77	0.8	0.32	10	3.20	37	7.25	85	28.25
E4	9	3.18	0.3	0.11	0	0	0	0	0	0

TABLE 2 Total variation (Δ) and standard deviation (σ) of transmitter frequency ν in Hz and reference voltage TX in Volt and linear gradient offsets G_x , G_y , G_z for the four experimental series.

simulations were performed using the FID-appliance (FID-A) [48], a simulation and data processing package for MRS. MEGA-PRESS difference spectra were generated, mirroring the sequence parameters of the phantom study with a finite-bandwidth editing pulse alternating between 1.9 ppm and 7.4 ppm at a field strength of 2.89 T with an acquisition bandwidth of 1250 Hz and two and four step phase cycling. Spectra were generated for GABA, Cr, NAA, Glu and Gln. The difference spectra were combined according to the concentrations used in the phantom experiments. The metabolite models simulated were derived from the work of Govindaraju *et al.* [10]. For GABA, there is a greater degree of uncertainty [16] and basis sets were generated for three distinct models that have been popularised in the literature by Govindaraju *et al.* [10], Kaiser *et al.* [13] and Near *et al.* [29], respectively. Simulated series were tested with all three GABA models to investigate any potential bias of our simulated data. For all simulated series, GABA concentrations were varied between 1 mM and 12 mM at 1 mM intervals and the combined spectra were analysed using our in-house code.

3 | RESULTS AND ANALYSIS

We initially assess the quality of the spectra and experimental calibration, before presenting the quantification results obtained and simulation results to help elucidate the experimental findings.

3.1 | Calibration and characterisation of spectra

The stability of the experimental setup over the course of an entire series of experiments was assessed by considering the variation of core parameters such as transmitter voltage and frequency, shim gradients and water suppression settings. Table 2 shows that a small adjustment made to the experimental procedure for E3 and E4 to avoid table movements between successive scans significantly reduced adjustments of transmitter frequency, reference voltage and linear gradient offsets G_x , G_y , G_z , although the results are within acceptable limits for all four experimental series. The difference spectra are shown in Appendix B.1.

Table 3 gives the linewidths of the water and NAA peaks obtained from Lorentzian peak fits, indicative values for the water suppression factors (WSF) and the SNR of the NAA peak. While the linewidths increase for the gel phantoms as expected, the width of the NAA peak in the difference spectrum remains around 4 Hz, suggesting good field homogeneity and frequency stability over the duration of the experiments. SNRs over 200 for NAA in the solution series and over 100 for the gel phantoms are also well within the acceptable range.

	H ₂ O WS OFF	H ₂ O EDIT ON	H ₂ O EDIT OFF	WSF	NAA EDIT OFF	NAA DIFF	SNR NAA
E1	2.40 ± 0.67	2.10 ± 0.54	2.11 ± 0.55	1840	1.19 ± 0.31	1.19 ± 0.31	209.6
E2	3.04 ± 1.44	2.82 ± 1.03	2.82 ± 1.03	1873	1.84 ± 0.33	1.85 ± 0.33	223.0
E3	2.98 ± 0.90	4.16 ± 2.02	4.16 ± 2.01	665	1.97 ± 0.25	1.98 ± 0.25	222.0
E4	11.16 ± 3.19	12.42 ± 9.01	12.54 ± 8.91	478	4.20 ± 1.48	4.25 ± 1.55	109.7

TABLE 3 Linewidths (mean/std over series in Hz) of H_2O peak in spectra acquired without water suppression (WS OFF) and the EDIT ON/OFF spectra with water suppression, NAA peak for EDIT ON and difference (DIFF) spectra with water suppression, as well as median water suppression factor (WSF) and SNR of NAA peak.

3.2 | Quantification

The quantification results obtained are summarised in Table 4. The results are reported as slope and intercept with 95% confidence intervals and R^2 value of a linear fit of the GABA-to-NAA amplitude ratios versus the actual concentration ratios. The amplitude ratios are also plotted against the known experimental concentration ratios in Figure 2a-Figure 2d. For the tools reporting peak areas rather than amplitudes (GANNET, LWFIT) the area ratio was multiplied by $\frac{3}{2}$ to account for the proton weightings of the GABA and NAA peaks but no other adjustments were made.

Figure 2b shows the GABA-to-NAA ratio for the pH calibrated solution series E1. The most notable feature of this series is that TARQUIN overestimates it by almost a factor of three. All other tools underestimate GABA to varying degrees from less than a third (0.279) of the actual ratio for GANNET (without adjustments) to around two thirds for LCM, JMRUI AQSES and LWFIT. The *R*² values of the linear fit also vary quite considerably between tools, with the highest correlation (99.5%) observed for LCM, followed closely by LWFIT, and the lowest (65%) for JMRUI. While intercepts of the linear fits for most tools are close to zero, both the AQSES and QUEST versions of JMRUI give a very large offset of almost 0.5, which suggests that some signal unrelated to GABA is erroneously assigned to GABA. As both fitting algorithms exhibit this problem, it can likely be attributed to the pre-processing or basis set.

Figure 2b shows the results for the intentionally miscalibrated solution series. Although R^2 drops only slightly, LCM performs significantly worse, under-estimating the GABA-to-NAA ratio at around 24% of the expected value. Given the pH dependent components of the NAA signal that will not be properly captured by the basis set spectra, it is unsurprising that basis set quantification methods perform worse. Surprisingly, however, both TARQUIN and JMRUI give significantly more accurate results for this series, with slopes close to 1 and higher R^2 than for the pH-calibrated series. For TARQUIN this may be a fortunate consequence of its previously observed tendency to overestimate GABA. However, this does not explain the significant improvement seen for JMRUI. The accuracy of the quantification for GANNET, while expected to be less sensitive to pH changes, also improves, although the substantial overlap in the confidence intervals of the parameters suggests that the improvement may not be significant. LWFIT also appears to be robust to the pH changes, with GABA gradient estimation exhibiting only a modest reduction, from 69% for E1 to 54% for E2, while maintaining a high R^2 of 0.97, as expected for a tool that only uses the 2 ppm NAA and 3 ppm GABA peaks, which are independent of pH.

Figure 2c shows the results for the pH-calibrated solutions with Glu and Gln. We observe underestimation of the GABA-to-NAA ratio at about 30% of the expected value for GANNET (without adjustments), more modest underestimation at around 60% for LCM and LWFIT, and a reduced degree of overestimation of the GABA-to-NAA ratio at 185% for TARQUIN, largely mirroring the results for E1. The most pronounced change is for JMRUI. While both fitting routines (QUEST and AQSES) resulted in underestimation of the GABA-to-NAA ratio in E1 at around 60%, both now overestimate the GABA-to-NAA ratio at 140%. The large changes could be due to systematic misidentification of

TABLE 4 Linear fit results for GABA-to-NAA ratios in all four experimental series. Reported are the gradient *a* and intercept *b* of the linear fit with 95% confidence intervals and the coefficient of determination R^2 (to three significant digits). Perfect quantification corresponds to a = 1, b = 0 and $R^2 = 1$. The percentage variation of reported amplitudes for NAA, whose actual concentration is constant, is also given. The large % variation of NAA amplitudes reported by TARQUIN for E1 is due to an unexplained rescaling of metabolite amplitudes between scans 10 and 11. The % variation of NAA for scans 1 – 10 is 6.09%. There is no simple explanation for the large variation reported by TARQUIN for E4.

Series	Analysis Tool	Gradient a with 95% Cl	Intercept b with 95% CI	R ²	NAA %
E1	TARQUIN	2.85 (2.41, 3.30)	–0.116 (–0.271,0.0526)	0.946	38.2
	JMRUI (AQSES)	0.665 (0.365,0.965)	0.475 (0.366,0.583)	0.684	3.86
	JMRUI (QUEST)	0.572 (0.297,0.847)	0.484 (0.385,0.584)	0.656	4.03
	LCM	0.625 (0.597,0.654)	0.0164 (0.00602,0.0267)	0.995	2.76
	GANNET 3	0.279 (0.209,0.349)	-0.0126 (-0.0381,0.0128)	0.875	3.39
	LWFIT	0.689 (0.61,0.768)	0.00856 (-0.0201,0.0373)	0.951	2.61
	LWFIT (RAW SMOOTH)	0.642 (0.582,0.701)	0.0188 (-0.0026,0.0402)	0.981	3.98
E2	TARQUIN	1.07 (0.869,1.27)	0.051 (-0.0261, 0.128)	0.903	2.87
	JMRUI (AQSES)	1.01 (0.730,1.29)	0.271 (0.164,0.378)	0.811	4.95
	JMRUI (QUEST)	1.06 (0.722,1.39)	0.279 (0.151,0.4071)	0.766	5.52
	LCM	0.239 (0.207,0.271)	0.0131 (0.000679,0.0254)	0.947	4.33
	GANNET 3	0.335 (0.304,0.366)	-0.0216 (-0.0333,-0.00977)	0.975	2.85
	LWFIT	0.535 (0.48,0.59)	-0.00467 (-0.0164,0.0257)	0.969	4.22
	LWFIT (RAW SMOOTH)	0.523 (0.402,0.644)	0.0328 (-0.0134,0.0790)	0.860	5.35
E3	TARQUIN	1.85 (1.72,1.97)	-0.0462 (-0.116,0.0237)	0.987	5.55
	JMRUI (AQSES)	1.40 (0.866,1.93)	0.208 (-0.0846,0.500)	0.712	3.59
	JMRUI (QUEST)	1.415 (1.04,1.79)	0.231 (0.0251,0.436)	0.836	3.31
	LCM	0.599 (0.552,0.645)	-0.0444 (-0.07,0.0187)	0.983	3.77
	GANNET 3	0.30 (0.256,0.343)	-0.00856 (-0.0334,0.0153)	0.945	3.31
	LWFIT	0.625 (0.562,0.687)	0.0299 (-0.00399,0.0637)	0.973	1.59
	LWFIT (RAW SMOOTH)	0.613 (0.548,0.678)	0.0406 (0.0053,0.0758)	0.969	2.60
E4	TARQUIN	1.77 (-0.544,4.08)	0.131 (–0.695,0.957)	0.368	42.1
	JMRUI (AQSES)	1.17 (0.154,2.18)	0.478 (0.116,0.840)	0.570	5.85
	JMRUI (QUEST)	1.02 (0.184,2.22)	0.521 (-0.0915,0.950)	0.417	6.11
	LCM	0.451 (0.217,0.685)	0.036 (-0.0476,0.12)	0.788	7.83
	GANNET 3	0.255 (0.168,0.342)	0.043 (0.0119,0.0742)	0.896	2.82
	LWFIT	0.585 (0.349,0.82)	-0.0325 (-0.0517,0.117)	0.860	2.29
	LWFIT (RAW SMOOTH)	0.607 (0.369,0.846)	0.0341 (–0.0512,0.119)	0.866	5.23



(c) E3 phantom series results.



FIGURE 2 GABA-to-NAA amplitude ratios reported by different tools vs. actual concentration ratio.

Glu and Gln in the difference spectrum, falsely attributing it to GABA or NAA. It is somewhat surprising to see larger changes for basis set methods (except LCM) as basis functions should be able to separate individual components better than peak-fitting approaches. The quality of the GABA-to-NAA fits is better across the board, with higher R^2 values and less variation in R^2 for different tools, ranging from R^2 of 0.7 to 0.8 for JMRUI to R^2 values over 0.95 for GANNET, LCM and LWFIT. The overall improvement in the quality of the linear fits is unlikely to be related to the introduction of Glu and Gln, but could be due to a small change in the experimental procedure resulting in fewer and smaller adjustments in transmitter settings and shim gradients between scans as illustrated in Table 2.

Figure 2d shows the results for the gel phantom series with Glu and Gln. The fit quality for the gel series is generally lower, as expected due to higher FWHM and lower SNR for tissue-mimicking gels, with the highest R^2 between 0.85 and 0.9 achieved for peak fitting and integration methods (GANNET, LWFIT). Among the basis set methods, LCM again outperforms TARQUIN and JMRUI in terms of linearity with R^2 around 0.8 vs. 0.4 to 0.6 for TARQUIN and JMRUI. TARQUIN again overestimates the GABA-to-NAA ratio at 177%, while LCM, GANNET and LWFIT underestimate the GABA-to-NAA ratio ranging from 26% to 59% of the actual ratio (without adjustments for editing efficiency). While the slope of the fit for JMRUI is close to 1, the large offset at around 0.5 and very low R^2 suggest that this is incidental and not indicative of accurate quantification.

3.3 | Simulation results

The experimental quantification results can be partially explained by simulations. Figure 3a shows the GABA-to-NAA amplitude ratio obtained by LWFIT vs. the actual GABA-to-NAA ratio for three series of simulated experiments. The blue line corresponding to the GABA-to-NAA ratios for MEGA-PRESS spectra generated assuming ideal excitation and refocusing pulses has a slope of 1.01, within 1% of the actual ratio, and an intercept of -0.05, which suggests a small negative offset, and an R^2 value of 1.00 (up to 3 significant digits). This suggests that the ratio of the GABA and NAA peak areas indeed scale linearly with the concentration ratio and the actual ratio is reproduced with very high accuracy by this method for the ideal spectra. However, for spectra simulated assuming finite-duration refocusing pulses with realistic pulse envelopes, the GABA-to-NAA ratio is only 84% of its actual value for the NAA-GABA-Cre series and 80% when Glu and Gln are added. There is also a marginal reduction in the negative offsets, while the linearity is maintained with $R^2 = 1.00$ throughout. This suggests that imperfect refocusing pulses may be one reason for the underestimation of GABA-to-NAA ratios.

Spatially resolved spectra further confirm this. Figure 3 shows the 3 ppm GABA peak for the edit-on, edit-off and difference spectra as a function of position on a 2D grid. The difference spectra near the boundary of the selected voxel indicate that the editing efficiency is significantly reduced. In an experimental setting the editing efficiency will be further reduced by imperfect slice profiles of the excitation pulses and other factors such as field inhomogeneity. Assuming these issues lead to a further reduction of the editing efficiency by a factor of 0.84 and 0.80, respectively, the simulations predict GABA-to-NAA ratios of 0.7 and 0.64 of the actual value for E1 and E3, respectively, which is very close to the values obtained by LCM and LWFI – 0.689 and 0.625 for LWFIT for E1 and E3, respectively. It is reasonable to assume greater underestimation for the intentionally miscalibrated series (E2) and the gel phantom series (E4), which is the case for LWFIT (0.535 and 0.585) and LCM (0.239 and 0.451). The much larger underestimation error for LCM is unsurprising as basis set algorithms are more susceptible to pH miscalibration, especially considering that NAA, our reference compound, has several minor pH-dependent peaks.

For TARQUIN we observe a reasonably consistent pattern of over-estimation in that the GABA-to-NAA ratio obtained for E1 (2.91) is larger than for E3 (1.85), which is in turn slightly larger than for E4 (1.77) and quite significantly larger than for E2 (1.07). The tendency of TARQUIN to overestimate the GABA-to-NAA ratio could be due to an internal calibration factor designed to adjust for underestimation due to experimental factors, in particular considering that TARQUIN appears to give the most accurate results for the intentionally miscalibrated series E2. The relatively poor performance of TARQUIN for MEGA-PRESS difference spectra could also be explained by its use of a pseudo-model for the difference spectra, which changes the expected phase of the 2.0 ppm NAA peak and fits GABA as a series of Lorentzians, unlike for non-edited spectra, for which it uses a fully simulated basis set.

The JMRUI results are more difficult to interpret. The GABA-to-NAA ratios are underestimated (0.665 and 0.572) for E1 but significantly overestimated (1.40 and 1.415) for E3. Combined with the low values of R^2 and the large constant offsets, this suggests that there are problems with the fitting or pre-processing of the spectra. The significant increase of the GABA-to-NAA ratio for E3 compared to E1 suggests that some of the Glu and Gln signal may be misidentified as GABA. The large offsets suggest problems with baseline fitting, or inaccuracies in the generated basis set.

The GANNET results also require further investigation. Although the R^2 values of the linear fits are comparable to other tools, ranging from 0.875 to 0.975, the degree of underestimation of the GABA-to-NAA ratios reported at 0.279, 0.335, 0.30 and 0.255 is surprisingly large. For the quantification of *in-vivo* data, Gannet attempts to correct for editing efficiency, explicitly assuming an efficiency of 50% for the 3 ppm GABA peak. This is somewhat lower than suggested by our simulation results, but might be attributed to different simulation parameters. However, even applying this



(a) GABA quantification results (LWFIT) for series of simulated spectra and linear fit parameters with 95% confidence intervals.



(b) Edit on, edit off and difference spectra for GABA (Near model). 2D simulations performed using FID-A on an 8×8 grid over a $5 \text{ cm} \times 5 \text{ cm}$ region with a target excitation volume of $2 \text{ cm} \times 2 \text{ cm}$ in the centre.

FIGURE 3 2D simulations indicate that shaped refocusing pulses tend to significantly reduce the GABA editing efficiency near the boundary of the voxel, resulting in underestimation of GABA.

factor to the GABA-to-NAA ratios reported in Table 4, an underestimation of the gradients is still observed. It is also surprising that the best estimation results are obtained for the intentionally miscalibrated series E2. Further investigation and comparison with LWFIT suggests that the main reason for the underestimation is the way the area of the GABA peak is computed by fitting a Gaussian using the default fitting routine. The results can be improved under certain conditions using GANNET's PhantomFit routine, details of which are discussed in Appendix D.

As different simulators use different models for GABA, e.g., LCM supports the Govindaraju and Kaiser model, while JMRUI uses its own modified version of the Govindaraju and Kaiser models, we also performed simulations to assess model dependence. GABA spectra simulated using the Govindaraju, Kaiser and Near model, shown in Figure 4



FIGURE 4 Simulated GABA spectra for the three models considered. Kaiser *et al.* and Near *et al.* models differ only in some couplings, whereas Govindaraju's model produces a significantly different spectrum.

	TARQUIN	JMRUI AQSES	JMRUI QUEST	LCM	GANNET	LWFIT
а	1.88 ± 0.734	1.06 ± 0.308	1.02 ± 0.346	0.479 ± 0.177	0.292 ± 0.0338	0.608 ± 0.0650
Ь	0.0068 ± 0.106	0.358 ± 0.139	0.379 ± 0.145	0.0053 ± 0.0346	0.0000659 ± 0.0291	0.0189 ± 0.0143
R ²	0.801 ± 0.291	0.694 ± 0.0991	0.669 ± 0.184	0.928 ± 0.096	0.922 ± 0.0458	0.943 ± 0.0554
%	39.0	29.1	34.1	37.0	11.6	10.7

TABLE 5 Mean and standard deviation of the linear fit parameters *a* (gradient) and *b* (intercept) and R^2 reported by different fitting tools and percentage variation (%) of the gradient over all series (std(*a*)/mean(*a*) × 100).

show clear differences especially in the structure of the 3 ppm GABA peak. For LWFIT the GABA-to-NAA ratios were 0.96, 1.00 and 1.00 for the Govindaraju, Kaiser and Near models, respectively, for the NAA/Cr/GABA series, and 0.93, 0.98 and 0.98, respectively, for the NAA/Cr/GABA/Glu/Gln series when ideal refocusing pulses were used in the simulations. While the difference between the models appears small in this analysis, especially between the Kaiser and Near models, numerical integration is relatively insensitive to the effects of coupling as *J*-evolved peak variations tend to average out over the integration window. The differences will be more significant for basis set fitting methods where the *J*-evolution is actively simulated. Generally, the Kaiser and the Near model appear to approximate the experimental spectra better than the older Govindaraju model.

4 | DISCUSSION

The results presented illustrate some of the shortcomings of the commonly used analysis software used for edited spectra. In terms of linearity, TARQUIN, GANNET and LCM perform well for the calibrated solution series but produce far more inconsistent results for the gel phantom spectra. This may be partly explained by the lower SNR and broader peaks of the gel spectra although the FWHM and SNR for the latter are still well within the acceptable limits suggested in a recent consensus paper [55]. As gel phantoms are tissue-mimicking [32] and the gel spectra resemble *in-vivo* spectra quite closely, for which the tools are designed, this is concerning.

With the exception of TARQUIN, which consistently overestimates GABA-to-NAA ratios, the tools studied tend

to underestimate GABA-to-NAA ratios to varying degrees. Simulations suggest realistic refocusing pulses reduce the editing efficiency and can lead to a reduction in reported GABA-to-NAA ratios of up to 20%. Including additional losses in experiments due to line broadening as a result of imperfect shimming and imperfect excitation pulses with realistic slice profiles could therefore explain the underestimation of GABA by many of the tools discussed.

LCM was one of the best performing tools for the calibrated solution series E1 and E3 with the highest R^2 values for the linear fits at 0.99. While still underestimating GABA-to-NAA ratios, the degree of underestimation is consistent with the losses expected based on simulation results, and only slightly worse than the results for LWFIT. However, LCM performs poorly for the intentionally miscalibrated solution series E2 and also struggles with the gel phantom spectra, resulting in greater underestimation and lower R^2 compared to the simple peak-integration method (LWFIT). This suggests that LCM is an excellent tool for well-calibrated spectra but may not be the best tool for more challenging spectroscopic environments. This conflicts somewhat with the findings of Marzola *et al.* [21], who report LCM as the best tool for low SNR spectra. However, the discrepancy may be due to LCM's baseline modelling procedure being more relevant for the lipid quantification study of Marzola *et al.* [21], its basis set fitting algorithm being particularly sensitive to pH miscalibration, and its baseline fitting algorithm performing especially poorly for gel phantoms.

Although JMRUI performed poorly in our study, the results present a useful insight into the effect of the fitting algorithm on quantification, independent of pre-processing, as both of the main fitting algorithms, QUEST and AQSES, use the same pre-processed spectra. While the differences in the results reported for both methods are negligible for E2 and E3, they are on the order of 10 - 15% for E1 and E4, which suggests a potential for diverging quantification results and the presence of a methodological bias, independent of pre-processing. Reproducibility of quantification results and the validity of cross-study comparisons are also limited due to JMRUI's significant user dependence with pre-processing, filtering and frequency and phase calibration all handled manually by the user, and basis set simulations subject to user-defined processing steps such as apodization. It is therefore possible that an experienced user could have improved its performance by carefully optimising each processing step.

A surprising result is that LWFIT, a simple peak integration method, appears to give the most consistent results across all series including for the gel spectra with R^2 values above 0.95 for the solution series and still among the highest for the gel spectra. While peak integration still underestimates the true concentration ratios in many situations, the estimates are more consistent and closer to the expected ratios than for any of the more sophisticated tools covered. While perhaps not a viable option for the quantification of complex spectra produced by standard spectroscopy sequences, simple methods such as this could be useful for the analysis of edited spectra, which are often simplified significantly. The results also suggests that peak fitting and extensive pre-processing steps such as filtering and baseline correction may actually be detrimental to the accuracy of GABA estimation, although more work is needed to properly substantiate this claim.

GANNET, the only non-basis set tool considered aside from our LWFIT code, performs worst in terms of the degree of underestimation of GABA-to-NAA ratios ranging from 0.335 of the actual value for E2 to 0.255 for E4 – without adjustments for editing efficiency – but has the lowest variation in GABA-to-NAA ratios (*m*) reported with $\sigma(m)/\text{ave}(m)$ of 11% (see Table 5). While its R^2 value for the calibrated solutions is lower than for LCM (0.875 vs. 0.995 for E1, 0.945 vs. 0.983 for E3), for example, it outperforms LCM for the gel phantoms (0.896 vs. 0.788 for E4) and miscalibrated solution series (0.975 vs. 0.947 for E2). The consistency of the underestimation is a testament to GANNET's robustness, but the accuracy of quantification in Gannet is dependent upon the correct modelling of edit efficiency. The underestimation of the GABA-to-NAA ratio – even after correcting for edit efficiency – can be partly attributed to the Gaussian model used by the standard fitting routine, and the results can be improved in some cases using GANNET's phantom fit routine.

Overall the results suggest that peak integration methods are more robust than basis set methods for edited MRS,

perhaps a consequence of the simplified spectral landscape, eliminating the requirement for basis set fitting. This in agreement with the observations of Marzola *et al.* [21], who also reported the robustness of these methods *in-vitro* but note the poor performance of integration methods *in-vivo*. However, due to the absence of a ground truth the accuracy of quantification is difficult to establish for *in-vivo* data.

5 | CONCLUSION

MEGA-PRESS spectra for carefully designed phantoms were analysed using different software tools commonly used in the community to quantify MRS data as well as our in-house code (LWFIT). Ratios of metabolite amplitudes, specifically GABA-to-NAA, were calculated and compared to the actual concentration ratios in the phantoms. Linear regression fits of the reported vs. actual ratios were performed to assess the overall quality and accuracy of the quantification. The results show that GABA-to-NAA ratios reported by different tools can differ substantially from each other and the ground truth of known concentration ratios for the phantoms. Compared to the ground truth, most tools assessed underestimate GABA-to-NAA ratios to varying degrees, although some such as TARQUIN consistently overestimated GABA-to-NAA ratios.

The tendency to underestimate GABA relative to NAA for the MEGA-PRESS implementation used in our study was consistent with simulations predicting underestimation due to reduced editing efficiency near voxel boundaries due to imperfect refocusing pulses and slice profiles. The accuracy of GABA estimation results therefore might be better for other implementations of the MEGA editing scheme, e.g., using adiabatic refocusing pulses. It would be desirable to repeat the calibrated phantom experiments for different sequences to identify the best-performing sequences and protocol parameters.

Of the tools considered (excluding our LWFIT code), LCM provides more accurate estimation results for wellresolved spectra but exhibits poor robustness to calibration errors and performs only moderately well for less wellresolved gel phantom spectra, which more closely resemble typical *in-vivo* spectra. GANNET appears to be the most robust tool considered, reporting the most consistent results across the experimental series, but it also exhibits the largest degree of underestimation of GABA-to-NAA ratios, some of which persists even when adjustments for editing efficiency are made. In some cases fitting results substantially improved by replacing the Gaussian fit by a model that better captured the pseudo-triplet nature of the GABA peak.

The systematic errors observed require further study to elucidate the precise nature of the variation and this work highlights the need for standardisation of existing methods and the development of new approaches to quantification of MRS data.

The methodological dependence of the quantification results observed also suggests that care must be taken when comparing results across studies, where analysis pipelines may differ and standardisation is desirable where possible. Furthermore, the methodological dependence is not restricted to instantaneous measurements of GABA, but is in fact found to propagate into reported changes in concentration, with tools presenting a range of GABA-to-NAA gradients. This finding is of great relevance to clinical studies, with our results suggesting that reported differences between normative and disease state GABA measurements will also be influenced by the choice of analysis method.

Acknowledgements

The authors would like to thank the Experimental MRI Center (EMRIC, Cardiff University) for the use of their LCModel license.

references

- Barkhuijsen H, De Beer R, Van Ormondt D. Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals. Journal of Magnetic Resonance (1969) 1987;73(3):553–557.
- [2] Beer RD, Ormondt DV. Analysis of NMR Data Using Time Domain Fitting Procedures Spectrum Analysis. In: Rudin M, editor. In-Vivo Magnetic Resonance Spectroscopy I: Probeheads and Radiofrequency Pulses Springer Berlin Heidelberg; 1992.p. 201–248.
- [3] Bhogal AA, Schür RR, Houtepen LC, van de Bank B, Boer VO, Marsman A, et al. ¹H-MRS processing parameters affect metabolite quantification: The urgent need for uniform and transparent standardization. NMR in Biomedicine 2017;30(11):1-9.
- [4] Bottomley PA. Spatial localization in NMR spectroscopy in vivo. Annals of the New York Academy of Sciences 1987;508(1):333-348.
- [5] Boy F, Evans CJ, Edden RAE, Lawrence AD, Singh KD, Husain M, et al. Dorsolateral prefrontal γ-aminobutyric acid in men predicts individual differences in rash impulsivity. Biological Psychiatry 2011;70(9):866–872.
- [6] Brix MK, Ersland L, Hugdahl K, Grüner R, Posserud MB, Hammar Å, et al. Brain MR spectroscopy in autism spectrum disorder—the GABA excitatory/inhibitory imbalance theory revisited. Frontiers in Human Neuroscience 2015;9:1–12.
- [7] Cousins MS, Roberts DC, de Wit H. GABAB receptor agonists for the treatment of drug addiction: a review of recent findings. Drug and alcohol dependence 2002;65(3):209–220.
- [8] Edden RAE, Puts NAJ, Harris AD, Barker PB, Evans CJ. Gannet: A batch-processing tool for the quantitative analysis of gamma-aminobutyric acid-edited MR spectroscopy spectra. Journal of Magnetic Resonance Imaging 2014;40(6):1445– 1452.
- [9] Ende G, Cackowski S, Van Eijk J, Sack M, Demirakca T, Kleindienst N, et al. Impulsivity and Aggression in Female BPD and ADHD Patients: Association with ACC Glutamate and GABA Concentrations. Neuropsychopharmacology 2016;41(2):410–418.
- [10] Govindaraju V, Young K, Maudsley AA. Proton NMR chemical shifts and coupling constants for brain metabolites. NMR in Biomedicine 2000;13(3):129–153.
- [11] Harris AD, Puts NAJ, Wijtenburg SA, Rowland LM, Mikkelsen M, Barker PB, et al. Normalizing data from GABA-edited MEGA-PRESS implementations at 3 Tesla. Magnetic Resonance Imaging 2017;42:8–15.
- [12] Jenkins C. New techniques for quantification of biomarkers and metabolites by magnetic resonance imaging and spectroscopy. PhD thesis, Swansea University; 2019.
- [13] Kaiser LG, Young K, Meyerhoff DJ, Mueller SG, Matson GB. A detailed analysis of localized J-difference GABA editing: Theoretical and experimental study at 4T. NMR in Biomedicine 2008;21(1):22–32.
- [14] Kanowski M, Kaufmann J, Braun J, Bernarding J, Tempelmann C. Quantitation of Simulated Short Echo Time ¹H Human Brain Spectra by LCModel and AMARES. Magnetic Resonance in Medicine 2004;51(5):904–912.
- [15] Ke Y, Cohen BM, Bang JY, Yang M, Renshaw PF. Assessment of GABA concentration in human brain using twodimensional proton magnetic resonance spectroscopy. Psychiatry Research - Neuroimaging 2000;100(3):169–178.
- [16] Kreis R, Bolliger CS. The need for updates of spin system parameters, illustrated for the case of γ-aminobutyric acid. NMR in Biomedicine 2012;25(12):1401–1403.
- [17] Lawson CL, Hanson RJ. Solving Least Squares Problems. SIAM; 1995.

- [18] Leal A, Vieira JP, Lopes R, Nunes RG, Gonçalves SI, Lopes da Silva F, et al. Dynamics of epileptic activity in a peculiar case of childhood absence epilepsy and correlation with thalamic levels of GABA. Epilepsy and Behavior Case Reports 2016;5:57–65.
- [19] Mandal PK. In vivo proton magnetic resonance spectroscopic signal processing for the absolute quantitation of brain metabolites. European Journal of Radiology 2012;81(4):e653–e664.
- [20] Marquardt DW. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. Journal of the Society for Industrial and Applied Mathematics 1963;11(2):431–441.
- [21] Marzola P, Mosconi E, Osorio Garcia MI, Sima DM, Van Huffel S, Fontanella M, et al. Different quantification algorithms may lead to different results: a comparison using proton MRS lipid signals. NMR in Biomedicine 2014;27(4):431–443.
- [22] McMahon RP, Chiappelli J, Krause BW, Kochunov P, Wijtenburg SA, Rowland LM, et al. Medial frontal GABA is lower in older schizophrenia: a MEGA-PRESS with macromolecule suppression study. Molecular Psychiatry 2015;21(2):198– 204.
- [23] Mescher M, Merkle H, Kirsch J, Garwood M, Gruetter R. Simultaneous in vivo spectral editing and water suppression. NMR Biomed 1998;11(6):266–72.
- [24] Mikkelsen M, Barker PB, Bhattacharyya PK, Brix MK, Buur PF, Cecil KM, et al. Big GABA: Edited MR spectroscopy at 24 research sites. NeuroImage 2017;159:32–45.
- [25] Mullins PG, McGonigle DJ, O'Gorman RL, Puts NAJ, Vidyasagar R, Evans CJ, et al. Current practice in the use of MEGA-PRESS spectroscopy for the detection of GABA. NeuroImage 2014;86:43–52.
- [26] Murdoch J, LCModel Basis Sets. Purdue University; 2011. http://purcell.healthsciences.purdue.edu/mrslab/basis_ sets.html, accessed 13th November 2018.
- [27] Myers JFM, Nutt DJ, Lingford-Hughes AR. γ-aminobutyric acid as a metabolite: Interpreting magnetic resonance spectroscopy experiments. Journal of Psychopharmacology 2016;30(5):422–427.
- [28] Naressi A, Couturier C, Devos JM, Janssen M, Mangeat C, De Beer R, et al. Java-based graphical user interface for the MRUI quantitation package. Magnetic Resonance Materials in Physics, Biology and Medicine 2001;12(2-3):141–152.
- [29] Near J, Evans CJ, Puts NAJ, Barker PB, Edden RAE. J-difference editing of gamma-aminobutyric acid (GABA): Simulated and experimental multiplet patterns. Magnetic Resonance in Medicine 2013;70(5):1183–1191.
- [30] Ogg RJ, Kingsley PB, Taylor JS. WET, a T1- and B1-insensitive water-suppression method for in vivo localized ¹H NMR spectroscopy. JMR B 1994;104:1–10.
- [31] Pan JW, Kuzniecky RI. Utility of magnetic resonance spectroscopic imaging for human epilepsy. Quant Imaging Med Surg 2015;5(2):313–322.
- [32] Portakal ZG, Shermer S, Jenkins C, Spezi E, Perrett T, Tuncel N, et al. Design and characterization of tissue-mimicking gel phantoms for diffusion kurtosis imaging. Medical physics 2018;45(6):2476–2485.
- [33] Poullet JB, Sima DM, Simonetti AW, De Neuter B, Vanhamme L, Lemmerling P, et al. An automated quantitation of short echo time MRS spectra in an open source software environment: AQSES. NMR Biomed 2007;20:493–504.
- [34] Poullet JB, Sima DM, Van Huffel S. MRS signal quantitation: A review of time- and frequency-domain methods. Journal of Magnetic Resonance 2008;195(2):134–144.
- [35] Provencher SW. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. Magnetic Resonance in Medicine 1993;30(6):672–679.

- [36] Provencher SW. Automatic quantitation of localized in vivo ¹H spectra with LCModel. NMR in Biomedicine 2001;14(4):260–264.
- [37] Puts N, Edden R. In vivo magnetic spectroscopy of GABA: a methodological review. Progress in Nuclear Magnetic Resonance Spectroscopy 2012;60:29–41.
- [38] Ratiney H, Coenradie Y, Cavassila S, Van Ormondt D, Graveron-Demilly D. Time-domain quantitation of ¹H short echo-time signals: Background accommodation. Magnetic Resonance Materials in Physics, Biology and Medicine 2004;16(6):284–296.
- [39] Ratiney H, Sdika M, Coenradie Y, Cavassila S, Ormondt Dv, Graveron-Demilly D. Time-domain semi-parametric estimation based on a metabolite basis set. NMR in Biomedicine 2005;18(1):1–13.
- [40] Reynolds G, Wilson M, Peet A, Arvanitis TN. An algorithm for the automated quantitation of metabolites in in vitro NMR signals. Magnetic Resonance in Medicine 2006;56(6):1211–1219.
- [41] Robertson CE, Ratai EM, Kanwisher N. Reduced GABAergic Action in the Autistic Brain. Current Biology 2016;26(1):80– 85.
- [42] Rowland L, Brooks WM, Lauriello J, Bustillo J, Mullins PG, Bedrick EJ. Reproducibility of ¹H-MRS measurements in schizophrenic patients. Magnetic Resonance in Medicine 2003;50(4):704–707.
- [43] Sanacora G, Mason GF, Rothman DL, Behar KL, Hyder F, Petroff OAC, et al. Reduced cortical γ-aminobutyric acid levels in depressed patients determined by proton magnetic resonance spectroscopy. Archives of General Psychiatry 1999;56(11):1043–1047.
- [44] Scott J, Underwood J, Garvey LJ, Mora-Peris B, Winston A. A comparison of two post-processing analysis methods to quantify cerebral metabolites measured via proton magnetic resonance spectroscopy in HIV disease. British Journal of Radiology 2016;89(1060).
- [45] Shermer SM, Jenkins C, Chandler M, Langbein FC, Magnetic resonance spectroscopy data for GABA quantification using MEGAPRESS pulse sequence. IEEE Dataport; 2019. http://dx.doi.org/10.21227/ak1d-3s20.
- [46] Sima DM, Van Huffel S. Separable nonlinear least squares fitting with linear bound constraints and its application in magnetic resonance spectroscopy data quantification. Journal of Computational and Applied Mathematics 2007;203(1):264–278.
- [47] Simpson MDC, Slater P, Deakin JFW, Royston MC, Skan WJ. Reduced GABA uptake sites in the temporal lobe in schizophrenia. Neuroscience Letters 1989;107(1-3):211–215.
- [48] 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. Magnetic Resonance in Medicine 2017;77(1):23–33.
- [49] Soltani N, Qiu H, Aleksic M, Glinka Y, Zhao F, Liu R, et al. GABA exerts protective and regenerative effects on islet beta cells and reverses diabetes. Proceedings of the National Academy of Sciences 2011;108(28):11692–11697.
- [50] Starčuk Z, Starčukov J, Štrbăk O, Graveron-Demilly D. Simulation of coupled-spin systems in the steady-state freeprecession acquisition mode for fast magnetic resonance (MR) spectroscopic imaging. Measurement Science and Technology 2009;20(10).
- [51] Starčuk Jr Z, Starčuková J. Quantum-mechanical simulations for in vivo MR spectroscopy: Principles and possibilities demonstrated with the program NMRScopeB. Analytical Biochemistry 2017;529:79–97.
- [52] Stefan D, Di Cesare F, Andrasescu A, Popa E, Lazariev A, Vescovo E, et al. Quantitation of magnetic resonance spectroscopy signals: The jMRUI software package. Measurement Science and Technology 2009;20(10).

- [53] Taylor SF, Tso IF. GABA abnormalities in schizophrenia: a methodological review of in vivo studies. Schizophrenia research 2015;167(1-3):84–90.
- [54] Waddell KW, Avison MJ, Joers JM, Gore JC. A practical guide to robust detection of GABA in human brain by J-difference spectroscopy at 3 T using a standard volume coil. Magnetic Resonance Imaging 2007;25(7):1032–1038.
- [55] Wilson M, Andronesi O, Barker PB, Bartha R, Bizzi A, Bolan PJ, et al. Methodological consensus on clinical proton MRS of the brain: Review and recommendations. Magnetic resonance in medicine 2019;82(2):527–550.
- [56] Wilson M, Reynolds G, Kauppinen RA, Arvanitis TN, Peet AC. A constrained least-squares approach to the automated quantitation of in vivo ¹H magnetic resonance spectroscopy data. Magnetic Resonance in Medicine 2011;65(1):1–12.

Supplementary Material

A | DETAILED DESCRIPTION OF PHANTOM PREPARATION

The solution series were prepared by dissolving the required concentrations of metabolites to remain constant in deionised water to create a base solution. To avoid variations in these background metabolites over a series of scans, a small amount of this base solution was removed using a syringe and GABA was added to prepare a solution with a GABA concentration of 1 mg/ml and the same concentration of the base metabolites. 290 ml of the GABA-free base solution were transferred to a round bottom glass flask, filling it to the neck. To vary the GABA concentration between scans, a small amount of the solution was removed from the flask with a syringe and replaced by the same volume of the concentrated GABA solution. The amount of solution replaced in each step varied, depending on the desired GABA concentration change, but was generally between 0.5 ml and 2 ml. This procedure was repeated several times, iterating GABA to a relatively high concentration. The same scan protocols were applied for each concentration step. This procedure limits experimental errors solely to the GABA concentration. Where stated, pH was monitored between scans and adjustments were made using a 36% hydrochloric acid (CAS-7647-01-0) solution and a 3.99% sodium hydroxide solution (CAS-1310-73-2) to maintain a pH of 7.2 ± 0.2 . The amounts of HCl or NaOH added after the initial pH calibration of the base solution were marginal (\leq 1 ml for the 290 ml flask) and any dilution of the overall solution was therefore deemed negligible. E2 was one exception, where pH was intentionally mis-calibrated to investigate the effect this had on the various quantification procedures. The final pH of E2 was measured to be 3.0 ± 0.2 .

For the gel series, the phantoms were made in advance of scanning. 800 ml of a base solution of Glu, Gln, NAA and Cr was prepared and divided into eight 100 ml portions. Different amounts of GABA were added to each solution and pH calibration was performed as above. Finally, 1 g of agar (CAS-9002-18-0)² was added as gelling agent. The mixtures were then heated to 90 °C to 100 °C while being stirred until the agar had fully dissolved. A small hole (< 2 mm diameter) was created in a spherical plastic mould using a plastic welding tool and the solutions were injected via a syringe. The arrangement was then allowed to cool over night. Once solidified, the gels were examined and the opening was sealed using a small amount of silicone sealant. Spherically-shaped phantoms were deemed the most suitable to reduce magnetic susceptibility-induced field inhomogeneity and minimise the spectral linewidth.

The accuracy of the scale used for mass measurements is 1 mg and the graduated cylinders used for the volume measurements have an accuracy of 1 ml. Assuming negligible errors in the manufacturer's stated molar masses, the uncertainty Δc in the molar concentration $c = \frac{m}{MV}$, where *m* is the mass in grams, *M* is the molar mass of the solute and *V* is the volume of the solvent, is $\frac{\Delta c}{c} = \sqrt{\left[\frac{\Delta m}{m}\right]^2 + \left[\frac{\Delta Y}{V}\right]^2}$. Assuming (generously) an uncertainty of 2% for the mass measurements and 1% for the volume measurement, the uncertainty in the concentrations of base metabolites is estimated to be less than 3%. The uncertainty in the amount of solution added or removed from the flask in each step via 5 ml syringes is slightly higher and could be up to 10%, assuming an uncertainty of 0.1 ml per 1 ml to account for slight misreadings of the plunger position and tiny air pockets within the syringe.



FIGURE 5 Difference spectra aligned to 2.01 ppm NAA peak with 3 ppm GABA peak shaded. Vertically offset spectra correspond to increasing concentrations of GABA for phantoms and different volunteers for *in-vivo* example.

B | SPECTROSCOPY DATA ACQUIRED

B.1 | Difference spectra

The difference spectra acquired for series E2, E3 and E4 shown in Figure 5, are consistent with the structure predicted by simulations and closely resemble the spectra predicted for the Kaiser/Near model. The expected pseudo-triplet structure, clearly visible for the well-calibrated solution series E3, is still discernible in the gel phantom spectra E4 but the lineshape is broader and more variable, while it is largely obscured for the intentionally miscalibrated series E2. Spectra for E1 are similar to E3 and have been omitted. A number of *in-vivo* spectra acquired using the same sequence and protocol on the same scanner ($2 \times 2 \times 2$ cm³ voxel in the prefrontal cortex of healthy volunteers) shows slightly broader peaks but a similar structure to our phantom spectra, especially our gel phantom spectra.

B.2 | Raw data vs. vendor-supplied dicoms

For the purpose of this paper, the combined channel and average spectra generated by vendor-supplied spectroscopy software in dicom format were used. Although this is advantageous for consistency and the spectra are generally suitable, some of spectra generated by the vendor-supplied software are reconstructed with relative phase errors, resulting in invalid difference spectra upon subtraction. Our own raw-data reconstruction was able to rectify this issue, but for compatibility with the various tools, the dicoms were preferred. All data included in the comparative analysis were screened for this problem to ensure valid difference spectra were included.

C | OVERVIEW OF ANALYSIS SOFTWARE

C.1 | LWFIT

Our in-house code, LWFIT, written in MATLAB, aligns the edit-off and difference spectra so that the main NAA peak is precisely at 2.01 ppm and then calculates the areas for the NAA (1.91 ppm to 2.11 ppm, difference), Cr (2.90 ppm to 3.10 ppm, edit-off) and GABA (2.90 ppm to 3.12 ppm, difference) peaks, as well the two main peaks associated with Glu and Gln, GLX1 (2.25 ppm to 2.45 ppm, difference) and GLX2 (3.65 ppm to 3.85 ppm, difference), using the real part of the spectra. This is done by numerical integration using piecewise-linear functions over the indicated fixed ppm ranges, selected to minimise contamination from other signals. Lorentzian, Gaussian and spline fitting and several baseline fitting and filtering methods were explored but initial testing indicated that numerical integration with minimal pre-processing yielded the more accurate estimation results compared to the ground truth. While more aggressive noise filters aesthetically improve the quality of the spectra, their application tends to increase the underestimation of GABA, especially for weak signals, suggesting that filtering eliminates some of the GABA signal present in the data [12]. For this work, only zero-filling (N = 4096 points) and a frequency shift to align the NAA peak in the spectrum to 2.01 ppm are performed. A Hanning filter of length 3500 is (optionally) applied to spectra computed directly from raw data (Raw Smooth version). No additional filters are applied to the coil-and-channel combined spectra produced by the vendor-supplied software. The results are reported as peak areas and ratios.

²Food-grade agar was sourced for this study

C.2 | TARQUIN

Totally Automatic Robust Quantitation in NMR (TARQUIN) [40, 56] is a cross platform, time domain basis set analysis tool. TARQUIN is an open source software package written in C++, complete with a GUI (Graphical User Interface) and a built-in NMR simulator. TARQUIN uses a Lawson-Hanson non-negative least-squares algorithm [17] with a basis set of pre-simulated spectra, which can be provided by the user. Residual water removal is performed using HSVD (Hankel Singular Value Decomposition) [1, 2] and automated phase and frequency correction is applied. TARQUIN does not perform a full MEGA-PRESS simulation, but rather models the expected signal and adjusts its phase correction procedures to accommodate the negative NAA peak. Frequency calibration is made relative to NAA. Results are reported as fit amplitudes. For this paper, TARQUIN version 4.3.11 with its internal MEGA-PRESS basis set was used.

C.3 | JMRUI

JMRUI [28, 52] is proprietary software distributed under its own license terms and made freely available to registered users for non-commercial use. It has a large suite of pre-processing, analysis and simulation options. Macromolecular baselines are fitted non-parametrically using penalised splines. For this study, two basis set methods were used to quantify the spectra. The first, quantitation based on semi-parametric quantum estimation (QUEST) [38, 39], is a time-domain fitting tool, which includes a semi-parametric approach to handle spurious signals resulting from macro molecules and lipids. The second method, automated quantitation of short echo time MRS spectra (AQSES) [33], uses a modified VARPRO variable projection NLLS algorithm [46], with imposes prior knowledge in the form of upper and lower bounds on the nonlinear parameters. For this study JMRUI version 6.0 beta was used. MEGA-PRESS basis sets were generated using NMR-SCOPE-B [50, 51] by defining a single MEGA-PRESS pulse sequence with sequence parameters matching our experimental protocol. Two basis sets for GABA/Cr/NAA and GABA/Cr/NAA/Glu/Gln were generated. Manual frequency and phase calibration and apodization were performed on the input spectra and fitting was applied using this common input. Results reported are the fit amplitudes.

C.4 | LCModel

LCModel (LCM) [35, 36] is a commercial, Linux-based closed-source MRS analysis tool that can be used with its internal *in-vitro* basis set or any arbitrary basis set specified by the user. The baseline signal resulting from macromolecules is established using spline fits. Fitting is attempted using a Gauss-Newton non-linear least-squares algorithm with Marquardt modification [20] with additional terms for T_2 , spectral shift and field inhomogeneities. For this paper version 6.3-1L and the Purcell lab basis set [26] recommended by LCM's creators were used.

C.5 | GANNET

GABA-MRS analysis tool (GANNET) [8] is an open-source automated MATLAB-based analysis tool, specifically designed for automated processing of 3 T GABA MEGA-PRESS data. GANNET consists of two separate modules, GAN-NETLoad and GANNETFit. The first module receives time-domain data, performs channel combination, adds line broadening, frequency and phase corrections, outlier rejection and time averaging. The output structure is then analysed by the fitting module. Fitting is performed using non-linear least-squares algorithms (lsqcurvefit and nlinfi) and is based on a five parameter Gaussian model to estimate the 3 ppm GABA signal in the difference spectrum, a six parameter Lorentzian model to estimate the 3 ppm Cr signal in the off spectrum and, if appropriate, a six parameter



FIGURE 6 Example of 3 ppm GABA peak for the calibrated series E1 (left) and intentionally miscalibrated series E2 (right) with GANNET Gaussian fit (red) and residuals, showing poor fit and large residuals for E1, where the triplet structure is clearly visible, but much better fit for E2, where the triplet structure is obscured due to line broadening.

Gaussian-Lorentzian model to fit the unsuppressed water signal. Quantitative results are reported as integral ratios of GABA to Cr and a concentration relative to water, NAA or Glx (Glu and Gln combined). While GANNET attempts to remove user interaction from the analysis procedure, it allows user modification of the code and adjustment of the pre-initialisation script is encouraged depending on setup. For this paper GANNET version 3.0 was used and the pre-initialisation script was adjusted to account for the fact that the phantoms spectra were acquired at room temperature.

D | GANNET FIT VS GANNET PHANTOM FIT

As noted in the discussion, GANNET's standard fitting routine underestimates GABA-to-NAA ratios more significantly than other tools, and a degree of underestimation persists even after adjusting for editing efficiency. For comparison, without adjustments for editing efficiency, LWFIT, a basic peak integration method, estimates GABA-to-NAA ratios at about 60% of the actual value, close to the value expected considering the editing efficiency suggested by our simulations, while the GANNET estimates are generally well below 50%, so still correspond to underestimation even when a correction factor of 2 is applied.

Underestimatation of GABA is more probable in view of Figure 6 showing that the pseudo-triplet structure of the GABA peak for a well-resolved phantom spectrum (E1) is poorly fit by a Gaussian, leading to a large residual, while the structure of the GABA peak in the experimental spectrum for the intentionally miscalibrated series E2 is obscured and the peak appears more Gaussian, resulting in a better fit for the Gaussian model. The intentional miscalibration has a similar effect to introducing field inhomogeneity, resulting in line broadening. To deal with non-Gaussian GABA peaks GANNET has an alternative fitting routine GannetFitPhantom, which attempts to fit pseudo-triplet GABA peaks by a triplet of Lorenzians. While our primary aim was to benchmark the routines commonly used for quantification of *in-vivo* spectra, we also quantified each of the spectra using both the standard (Gaussian) fitting routine and the phantom-fitting routine.

Figure 7 shows the phantom fit routine indeed performs somewhat better for series E1 and E3, while the regular Gaussian fitting routine performs slightly better for E2, as expected. The differences are not very large, however. Sur-



(c) Series E3

24

(d) Series E4

FIGURE 7 Comparison of GannetPhantomFit and the regular GannetFit routine

prisingly, the phantom fit routine appears to have a more substantial advantage for the tissue-mimicking gel phantoms, for which we expected the Gaussian fit routine typically used for *in-vivo* spectra to have an edge.