Assessment of measurement precision in single-voxel spectroscopy at 7 T: Toward minimal detectable changes of metabolite concentrations in the human brain in vivo

Layla Tabea Riemann1 | Christoph Stefan Aigner1 | Stephen L. R. Ellison2 | Rüdiger Brühl1 | Ralf Mekle3 | Sebastian Schmitter1 | Oliver Speck4,5 | Georg Rose5,6 | Bernd Ittermann1 | Ariane Fillmer1

1Physikalisch-Technische Bundesanstalt, Braunschweig und Berlin, Germany
2LGC Limited, Teddington Middx, United Kingdom
3Center for Stroke Research Berlin, Charité–Universitätsmedizin Berlin, Berlin, Germany
4Biomedical Magnetic Resonance, Otto-von-Guericke University, Magdeburg, Germany
5Research Campus STIMULATE, Magdeburg, Germany
6Institut für Medizintechnik, Otto-von-Guericke University, Magdeburg, Germany

Purpose: To introduce a study design and statistical analysis framework to assess the repeatability, reproducibility, and minimal detectable changes (MDCs) of metabolite concentrations determined by in vivo MRS.

Methods: An unbalanced nested study design was chosen to acquire in vivo MRS data within different repeatability and reproducibility scenarios. A spin-echo, full-intensity acquired localized (SPECIAL) sequence was employed at 7 T utilizing three different inversion pulses: a hyperbolic secant (HS), a gradient offset independent adiabaticity (GOIA), and a wideband, uniform rate, smooth truncation (WURST) pulse. Metabolite concentrations, Cramér-Rao lower bounds (CRLBs) and coefficients of variation (CVs) were calculated. Both Bland-Altman analysis and a restricted maximum-likelihood estimation (REML) analysis were performed to estimate the different variance contributions of the repeatability and reproducibility of the measured concentration. A Bland-Altman analysis of the spectral shape was performed to assess the variance of the spectral shape, independent of quantification model influences.

Results: For the used setup, minimal detectable changes of brain metabolite concentrations were found to be between 0.40 µmol/g and 2.23 µmol/g. CRLBs account for only 16 % to 74 % of the total variance of the metabolite concentrations. The application of gradient-modulated inversion pulses in SPECIAL...
led to slightly improved repeatability, but overall reproducibility appeared to be limited by differences in positioning, calibration, and other day-to-day variations throughout different sessions.

**Conclusion:** A framework is introduced to estimate the precision of metabolite concentrations obtained by MRS in vivo, and the minimal detectable changes for 13 metabolite concentrations measured at 7 T using SPECIAL are obtained.

**KEYWORDS**
CRLBs, measurement precision, minimal detectable change, MR spectroscopy, reproducibility/repeatability, SPECIAL

1 | INTRODUCTION

In vivo $^1$H MRS allows the non-invasive detection of metabolic changes in various organs. In the brain, such changes are frequently associated with diseases of the central nervous system. In certain diseases, such as cancer, it is sufficient to obtain the ratios of the metabolites of interest, as they already differ significantly from healthy tissue. However, in neurodegenerative diseases, such as Alzheimer’s or Parkinson’s disease, or psychological disorders like schizophrenia, in which the metabolite differences between healthy and diseased tissues are more subtle, it is crucial to obtain absolute metabolite concentrations.

The use of ultrahigh field (UHF) strength offers increased SNR and enhanced spectral dispersion, which enables an improved distinction between overlapping metabolites.

Despite the potential of UHF-MRS to measure an increased number of metabolites more reliably, and thus to examine disease-related changes of the concentration levels of metabolites, so far it remains mostly a research tool with a focus on comparisons between patient and control cohorts, usually without the option for individual diagnosis based on MRS. Arguably, the most important reasons for this are: (1) the ranges of metabolite concentrations in healthy controls and patients often exhibit substantial overlap, and (2) no physically proven measurement uncertainties of concentrations obtained by MRS are available.

In other experimental areas, SDs and measurement uncertainties are determined by repeating the same measurement several times, and the result is ideally compared with a ground truth. Although phantoms with known concentrations can be used for repeated MRS measurements and may serve as ground truth for in vitro measurements, they still fail to mimic some technical aspects of the acquisition, the complex interplay between metabolite concentrations measured, tissue structure and partial-volume effects, and differences in the microenvironment of biological tissues. Moreover, in a clinical setting, it is usually impossible to repeat high-quality in vivo MRS measurements to estimate the SD properly. Hence, Cramér-Rao lower bounds (CRLBs), as suggested by Cassavila et al, are commonly used to estimate the reliability of measured metabolite concentrations.

To minimize CRLBs, which are negatively impacted by a reduced SNR due to $T_2$ relaxation and complex frequency signatures caused by long J-coupling evolution, short TEs ($\leq 10$ ms) are desirable to acquire spectra in single-voxel spectroscopy (SVS) at UHF. This facilitates the quantification of metabolites such as glutamate, glutamate, and $\gamma$-aminobutyric acid. Furthermore, the RF pulses used for localization should ideally be insensitive to $B_1^+$ inhomogeneities, as well as $B_0$ to ensure proper localization. Additionally, they should yield a small chemical shift displacement (CSD) to ensure that signals from different metabolites originate from the same location. Furthermore, low pulse energy and peak power are desirable to avoid limitations related to the specific absorption rate (SAR) or hardware capabilities.

Short TEs and high SNR in SVS can be achieved using the spin-echo, full-intensity acquired localized (SPECIAL)$^{17,18}$ spectroscopy sequence. Furthermore, a comparably low CSD and low sensitivity to $B_1^+$ is met by the adiabatic inversion pulse in SPECIAL. However, hyperbolic secant (HS)$^{20}$ pulses, as proposed for the pre-excitation inversion in the original implementation of SPECIAL using a surface coil, which also have been used in several subsequent studies with this sequence, require a relatively high peak power to reach adiabaticity compared with other adiabatic pulse types. In UHF applications in deeper brain structures, such as the hippocampus, this may lead to either the requirement of increasing the pulse duration resulting in a reduced bandwidth (BW) and hence a larger CSD - or reaching hardware limitations of available peak power - resulting in loss of adiabaticity - , and hence an incomplete inversion and unwanted signal loss.

More recently, gradient-modulated pulses, such as the gradient offset independent adiabaticity (GOIA)$^{27}$ pulse and the wideband, uniform-rate, smooth truncation (WURST)$^{28}$ pulse, were used in MRS studies. These pulses provide a substantially decreased CSD and sharper pulse profile compared with an HS pulse$^{29,30}$ of equal
duration and total pulse energy, while allowing to reduce peak power requirements for the same inversion BW. Since peak power is often a limiting factor, gradient-modulated pulses are a commonly used alternative. Furthermore, it can be expected that, in combination with slight imperfections of positioning and $B_0$ shimming, the decreased CSD and the sharper pulse profile exhibit a positive effect on the measurement precision.

Nevertheless, even if all the previously mentioned conditions are met and the CRLBs are minimized for a certain in vivo measurement, the CRLBs still fail to provide information on the precision of metabolite concentrations measured with a given experimental setup.\textsuperscript{31} The present work, therefore, aims to assess the repeatability and the reproducibility of in vivo metabolite concentrations obtained by SPECIAL MRS at 7 T, as well as the impact of different adiabatic inversion pulses within SPECIAL thereupon. The SPECIAL sequence using a nonadiabatic refocusing pulse was chosen here for its ability to yield a very short TE ($\leq 10$ ms). Results are then used to derive the measurement precision for concentrations of 13 metabolites in the human brain, unlike CRLBs, accounting not only for the lowest possible bound of the SDs of the model fit but also for instrumental and operational influences on the spectral data. Moreover, the minimal detectable change (MDC)\textsuperscript{32} of 13 metabolites in vivo for the used experimental setup is determined.

### 2 | METHODS

#### 2.1 | Inversion pulse implementation

An HS, GOIA, and WURST pulse was designed\textsuperscript{20,27,28} in MATLAB (The MathWorks, Natick, MA) to achieve identical pulse duration, inversion slice thickness, and pulse energy. Bloch simulations with varying $B^*_1$ and $B_0$ were performed.\textsuperscript{33} The three different pulses were then incorporated into the SPECIAL sequence, resulting in three different SPECIAL versions with otherwise identical scan parameters and timings. These variants will be referred to as HS-SPECIAL, GOIA-SPECIAL, and WURST-SPECIAL, respectively, throughout this paper. The resulting pulse sequence scheme is shown in Figure 1A.

#### 2.2 | MR protocol and data acquisition

All experiments were performed on a 7T scanner (MAGNETOM 7T, Siemens Healthineers, Erlangen, Germany) using a head coil with a birdcage transmitter and 32 receive channels (NOVA Medical Inc., Wilmington, USA). Phantom measurements were performed to assess the performance of the three SPECIAL variants without biological or physiological noise, as described in Supporting Information Figure S1 and Table S1.

### 2.3 | In vivo experiments

Nine healthy volunteers (aged 39 $\pm$ 13, 1:7:1 male:female:nonbinary) were scanned after giving written informed consent according to local ethical regulations, to assess the impact of the different adiabatic pulses on the variance of in vivo measurements. To this end, an unbalanced nested study design, as shown in Figure 1C, was chosen. Each volunteer was scanned in two sessions on two different days approximately one week apart (6 to 8 days). Both sessions consisted of two measurements (M1-M4) of SPECIAL acquisitions with the HS, GOIA, and WURST adiabatic inversion pulses, each. During session one, the volunteer was repositioned between M1 and M2, whereas in session two, M3 and M4 were acquired without repositioning in between. As the ethical regulations specify a maximum time of 90 minutes per scan block, the repeatability measurements in session two were split into two scan blocks with two SPECIAL versions, e.g., HS-SPECIAL and GOIA-SPECIAL, investigated in the first scan block, and the third SPECIAL version, e.g., WURST-SPECIAL, which was examined in the second one. Note that the subject was not repositioned between the repeatability measurements of the same SPECIAL version. A schematic overview of the different measurements, scan blocks, and sessions is shown in Figure 1C,D. The order of the SPECIAL versions within the different scan blocks was cyclically permuted among the different volunteers to ensure that the performance of the pulses is not biased due to the acquisition time point within the protocol, such as due to increased likelihood of volunteer movement toward the end of each scan block. With this design, it was possible to distinguish among three scenarios: (1) the repeatability\textsuperscript{34} ($R_o$), which refers to two consecutive measurements without repositioning the subject; (2) the reproducibility\textsuperscript{34} between two measurements performed on the same day, including repositioning and new calibration ($R_{1,M}$ for minutes in-between); and (3) the reproducibility between two measurements approximately one week apart ($R_{1,W}$ for week in-between). If the reproducibility scenarios could not be assessed individually, the index was extended by a ‘c’ for combined.

The protocol within each scan block was as follows: The MP2RAGE\textsuperscript{35} images acquired with the following parameters were used for voxel positioning and tissue segmentation: TE = 2.51 ms, TR = 5000 ms, TI = 900 ms, isotropic voxel size = 0.75 mm. The volume of interest (VOI) was placed in the posterior cingulate cortex (PCC) in the middle between both hemispheres and was angulated in the
sagittal plane so that its lower edge coincided with the virtual line between the corpus callosum and the outer end of the parieto-occipital fissure, as illustrated in Figure 1B. A voxel-based $B_0^+$ adjustment was performed by varying the pulse voltage and fitting the resulting amplitudes of the water peak to determine the voltage required for a 90° flip angle. First- and second-order $B_0$ shims settings were optimized for the VOI using a $B_0$ map (3 mm isotropic resolution, TE1 = 6.02 ms, TE2 = 7.04 ms, TR = 620 ms) and a MATLAB-based shim tool. Single-voxel spectra were
2.4 Spectral postprocessing

Spectra were post-processed with an in-house MATLAB tool, including the summation of the even and odd transient pairs, which were acquired with a 180° phase shift in the receive phase, to obtain the full localization. Then, weighted and phase-corrected coil-element combination, frequency correction based on the N-acetylaspartate (NAA) peak at approximately 2 ppm, and averaging were performed. Spectral quality was assessed for each SPECIAL version for all subjects, both qualitatively by visual inspection, and quantitatively by calculating the width and the SNR of the unsuppressed water line.

2.5 Metabolite quantification

The data sets were quantitatively analyzed using LCModel in the range of 0.2 to 4.2 ppm. A basis set for LCModel fitting containing signatures of alanine, aspartate (Asp), ascorbate, the sum of glycophosphocholine and phosphocholine (total choline - tCho), the sum of creatine and phosphocreatine (total creatine - tCr), γ-amino- butyric acid (GABA), glucose, glutamine (Gln), glutamate (Glu), glutathione (GSH), myo-inositol (Ins), lactate (Lac), NAA, N-acetylaspartylglutamate (NAAG), phosphoethanolamine (PE), scyllo-inositol, and taurine (Tau) was simulated in Vespa. The macromolecules were modeled in one basis function derived in-house from metabolite-nulled in vivo acquisitions in healthy volunteers using the SPECIAL sequence, as recommended by a recent consensus paper. The water signal was used as an internal standard to calculate concentration values.

2.6 Segmentation and tissue fraction correction

To compare the measured metabolite concentrations within each subject and among all subjects, the MP2RAGE images for every session and every volunteer were segmented into cerebrospinal fluid (CSF), gray matter (GM), and white matter (WM) with SPM12. Then, an in-house-written python tool was used to determine the GM, WM, and CSF fraction for the voxel. The LCModel output concentration was corrected for each voxel (c$_{i,j,m}$) to take the volunteer-specific and session-specific CSF fraction as well as relaxation processes into account:

\[ c_{i,j,m}^* = \frac{c_{i,j,m}}{\exp\left(-\frac{\text{TE}}{T_{2,m}}\right) \left(1 - \exp\left(-\frac{\text{TR}}{T_{1,m}}\right)\right) \left(1 - f_{\text{CSF},i,j}\right)} \]

where $c_{i,j,m}$ describes the concentration of the metabolite $m$ for volunteer $i$ and session $j$; $T_{1,m}$ and $T_{2,m}$ indicate the metabolite-specific relaxation times; and $f_{\text{CSF},i,j}$ indicates the CSF fraction for the volunteer $i$ and session $j$. The correction of the
relaxation times assumes that most of the voxel contains GM. Therefore, only the GM relaxation times from the investigated brain region are considered. However, GM and WM tissue fractions are considered to take $T_2$ effects of tissue water into account, which was used to obtain the attenuation factor for the water scaling performed in LCModel. $T_2$ relaxation times of 45 ms and 37 ms for GM and WM were used, respectively, leading to a water attenuation factor of 0.8111.

Voxel-position reproducibility was assessed by calculating the CSF, GM, and WM fraction for all four sessions and by determining the intra-subject coefficient of variation (CV) for the CSF fraction. In addition, an in-house-written python tool was used to determine the voxel overlap between two sessions by co-registration of two MP2RAGE images. A flow chart describing all postprocessing steps, the resulting data, as well as the resulting analysis can be found in Figure 1E.

### 2.7 Statistical analysis

Intra-subject CVs were calculated for each inversion RF pulse type, subject, and metabolite to assess the test-retest reproducibility quantitively. Statistical differences of the paired mean for each subject between the three SPECIAL variants were determined for metabolite concentrations, CRLBs, and CVs by a non-parametric Wilcoxon signed-rank test. Due to Bonferroni correction, the significance level of $p < .05$ was shifted to $p < .001$.

A summary and short explanation of the differently obtained SDs, which will be explained subsequently, can be found in Table 1.

Bland-Altman plots of the spectral shape were calculated as follows: First, the real part of the compared spectra was normalized to the intensity of the NAA peak. Then, to generate the $y$-value of one data point, $BA_i$, for subject $i$ in the Bland-Altman plots, the absolute of the

| | Index symbol | Explanation |
|---|-------------|-------------|
| Method (upper index) | S | SD of the Bland-Altman plots derived by the analysis of the spectral shape as described by Equations (2) and (3) |
| | BA | SD of the Bland-Altman plots for the corrected metabolite concentrations, $c_m^*$ |
| | REML | SD obtained by REML analysis from the corrected metabolite concentrations, $c_m^*$ |
| Scenario (lower index) | $R_0$ | Repeatability scenario: investigation of two measurements performed consecutively without any repositioning or recalibration within one scan block |
| | $R_{1,M}$ | Reproducibility scenario “minutes between measurements”: investigation of two measurements performed with repositioning and recalibration within one session |
| | $R_{1,W}$ | Reproducibility scenario “week between measurements”: investigation of two measurements performed in two sessions approximately one week apart |
| | $R_{1,Mc}$ | $= \sqrt{(\sigma_{R_0})^2 + (\sigma_{R_{1,M}})^2}$, combined SD of the $R_0$ and $R_{1,M}$ scenarios, the SD effectively observed in the $R_{1,Mc}$ scenario |
| | $R_{1,Wc}$ | $= \sqrt{(\sigma_{R_0})^2 + (\sigma_{R_{1,M}})^2 + (\sigma_{R_{1,W}})^2}$, combined SD of the $R_0$, $R_{1,M}$ and $R_{1,W}$ scenarios, the SD effectively observed in the $R_{1,Wc}$ scenario |

Note: The SDs were calculated for different scenarios and using different methods. The symbols for the SD, $\sigma_{method/scenario}$, follow the same pattern, with the upper index indicating the method used to determine the respective SD, and the lower index describing the scenario, for which the respective SD was calculated. The different methods and scenarios can be found in this table. Note, that ‘c’ in the scenario description indicates that the variance components of the scenario are combined and not considered individually. For the Bland-Altman plots of both the spectral shape and the concentrations, it is not possible to assess the components separately, only combined.
spectral intensity in the real part of the compared spectra, $|x(f)|_{M^1}$, was subtracted for each frequency $f$, and the integral within the frequency range from $f_{\text{min}} = 0.8$ ppm to $f_{\text{max}} = 4.2$ ppm was taken, as follows:

$$BA_{i,x} = \int_{f_{\text{min}}}^{f_{\text{max}}} |x(f)|_{M_{\text{a}}} - |x(f)|_{M_{\text{b}}} \, df.$$  

(2)

The x-value $BA_{i,x}$ was calculated as the integral of the absolute of the averaged real parts of the compared spectra, $|x(f)|$, over the same frequency range, as follows:

$$BA_{i,x} = \int_{f_{\text{min}}}^{f_{\text{max}}} |x(f)| \, df.$$  

(3)

The SDs $\sigma^x_{R_0}$, $\sigma^x_{R_{1,Mc}}$, and $\sigma^x_{R_{1,Wc}}$ for the points in the Bland–Altman plots give a measure for the precision of the spectral shape within the scenarios $R_0$, $R_{1,Mc}$, and $R_{1,Wc}$, respectively. The reproducibility SDs $\sigma^x_{R_{1,Mc}}$ and $\sigma^x_{R_{1,Wc}}$ could only be derived as a combined effect, however, as the data quality did not permit a robust separation of these contributions. The SDs for the metabolite concentrations for each scenario are also determined by a Bland–Altman analysis (REML), in which the difference between the concentrations obtained from two measurements was plotted over the arithmetic mean of them.

Furthermore, to quantify the measurement precision of individual metabolite concentrations obtained with the different SPECIAL variants, variance components of the metabolite concentrations were extracted separately for each metabolite/pulse combination using a restricted maximum likelihood estimation (REML) analysis, carried out in R version 3.6.3 using the nlme package. The statistical model used for the variance component extraction was:

$$C^\ast_m \sim \mu_m + S + P + \delta^\ast_{R_0} + \delta^\ast_{R_{1,Mc}} + \delta^\ast_{R_{1,Wc}},$$  

(4)

where $C^\ast_m$ is the relaxation-corrected concentration of the metabolite $m$; $\mu_m$ is the general mean of the concentration of metabolite $m$ for each inversion RF pulse type; $S$ is the subject effect; $P$ is the effect of the particular inversion pulse (HS, GOIA, or WURST); the three $\delta^\ast_{R_k}$ terms are random between-session effects, each with zero mean and variance $(\sigma^\ast_{R_k})^2$; $\delta^\ast_{R_0}$ is a residual error reflecting the measurement precision of back-to-back measurements, and hence the repeatability scenario $R_0$; $\delta^\ast_{R_{1,Mc}}$ is introduced through slight differences in positioning and calibration; $\delta^\ast_{R_{1,Wc}}$ reflects the combined measurement error of the reproducibility scenario $R_{1,Wc}$ after repositioning; $\delta^\ast_{R_{1,Mc}}$ is introduced through additional, e.g., physiological, random effects between two sessions a week apart; and $\delta^\ast_{R_{1,Wc}}$ reflects the combined measurement error of the reproducibility scenario $R_{1,Wc}$ after one week. Pulse and subject are considered as fixed effects for this analysis; it is assumed that subject and pulse effect comparisons are usually the target measure of a study, whereas the variances only need to be taken into account while interpreting the certainty of the results. All variances were assumed constant for each metabolite/pulse combination; the REML fit additionally assumed normality for the random effects and restricted the calculated SDs to the range of natural numbers. Because each variance estimate in a nested design has some impact on the next level ‘up’, the analysis and interpretation will be restricted to the within-group SDs of the repeatability $\sigma^\ast_{R_0}$, which are not affected by the other estimates, and the combined effect of all variances $(\sigma^\ast_{R_{1,Wc}})^2 = (\sigma^\ast_{R_0})^2 + (\sigma^\ast_{R_{1,Mc}})^2 + (\sigma^\ast_{R_{1,Wc}})^2$ rather than the individual contributions. In the absence of clear differences rooted in the application of the different adiabatic pulses, both SDs were averaged over the three pulses to assess the measurement precision of the different metabolite concentrations. The resulting SDs were then correlated to the CRLBs, as well as to $\sigma^\ast_{BA}$. The MDC was calculated from the REML analysis, using the standard error of measurement, SEM, as follows:

$$ \text{MDC} = 1.96 \cdot \sqrt{3} \cdot \text{SEM},$$  

(5)

$$ \text{SEM} = \sqrt{\left(\sigma^\ast_{\text{REML}}(\text{HS})\right)^2 + \left(\sigma^\ast_{\text{REML}}(\text{GOIA})\right)^2 + \left(\sigma^\ast_{\text{REML}}(\text{WURST})\right)^2}. $$  

(6)

3 | RESULTS

3.1 Inversion pulse implementation

Details of the different pulse parameters can be found in Table 2. The magnitudes, phases, and slice-selective gradients of the three pulses are shown in Figure 2A-C. The pulse duration, inversion slice thickness, and total pulse power were fixed and chosen such that the following conditions were met for all three pulses: (1) fulfilled adiabatic condition with reasonable safety margin while not reaching peak power limitations (shown in Figure 2H-J); and (2) a minimal inversion BW of 1.2 kHz.

Bloch simulations (Figure 2D-J) demonstrate that the BW of the gradient-modulated pulses is about 10 times higher than the BW of the HS pulse if pulse duration, FWHM of the slice thickness, and energy are fixed. This leads to a substantial reduction in CSD (Figure 2E-G), whereas the maximum RF amplitude is reduced by 33 % (Figure 2A). Scaled with the transmitter reference voltage calibrated in the in vivo measurements, this resulted in a difference in peak voltage of about 100 V between HS and gradient-modulated pulses, whereas the CSD of the gradient-modulated pulses is reduced by 90 % compared with HS (Table 2).
3.2 In vivo measurements

The voxel overlap between different measurements across all volunteers and all six possibilities was greater than 81% in all cases, as indicated in Table 3 and Figure 3. The intra-subject CV for the CSF fraction was 6.6 ± 4.9%.

Table 2 also lists the spectral quality parameters averaged over all volunteers and all 36 spectra for each SPECIAL version. Width and SNR of the water peak did not differ significantly among the three different inversion pulses in SPECIAL.

The spectral quality of spectra obtained with all three SPECIAL versions was high, and hardly any differences were discernable by visual inspection (Figure 4A). The Bland-Altman plots of replicate differences against the mean value (Figure 4B) revealed that in the R0 scenario (repeatability) the gradient-modulated pulses GOIA and WURST gave a smaller dispersion of the spectral-shape differences compared to the HS pulse. This difference in dispersion vanishes, however, when the reproducibility scenarios R1Mc and R1Wc were considered: HS-SPECIAL and WURST-SPECIAL were on a par, with GOIA-SPECIAL trailing slightly behind. Comparing the different scenarios, the order of the SDs was $\sigma_{R0}^S < \sigma_{R1Mc}^S < \sigma_{R1Wc}^S$. The Bland-Altman plots of the concentrations are depicted in Supporting Information Figures S2-S15.

Similar concentration variations were obtained for most of the quantified metabolites using the different SPECIAL versions, as shown in Figure 5A-B. However, the results from GOIA-SPECIAL and WURST-SPECIAL exhibited a higher concentration for tCr and Glu (both $p < .001$). For HS-SPECIAL, 14 individual concentrations (of the 468 values: 13 metabolites × 4 sessions × 9 volunteers) had to be discarded because they could not be quantified by LCModel, but only five and four metabolite concentrations for GOIA and WURST-SPECIAL, respectively. The CRLBs of Asp, NAA, and tCr were significantly higher for the measurement with HS-SPECIAL compared to the sequence variants using GOIA and WURST pulses, as shown in Figure 5B. There are no significant differences, both in concentration and CRLBs, between GOIA-SPECIAL and WURST-SPECIAL. HS-SPECIAL exhibited the highest averaged intra-subject CV for most of the metabolites, except for GABA and Lac, as displayed in Figure 5C. Correlation plots for the repeatability scenario for Glu, NAA, tCho, and tCr can be found in the Supporting Information Figure S16.

3.3 Precision evaluation

The pulse-wise $\sigma_{REML}$ for every metabolite for both the $K_0$ as well as the $K_{1,Wc}$ scenario are depicted in Figure 6A. Neither the SDs obtained from the repeatability measurements nor from the reproducibility measurements showed a consistent trend, favoring one of the investigated pulses. The individual results of the REML analysis can be found in Supporting Table S2. The MDCs, which were determined for every metabolite and the given setup, are depicted in Table 4 and Figure 6B.

The correlation between CRLBs and $\sigma_{REML}$, between $\sigma_{RA}$ and $\sigma_{REML}$, and between CRLBs and $\sigma_{RA}$, averaged over all pulses, are shown in Figure 6C-E. The lowest $R^2 = 0.809$ is found for the $K_{1,Wc}$ correlation between CRLBs and $\sigma_{RA}$, whereas the highest $R^2 = 0.99$ is found for the $K_0$ correlation between $\sigma_{RA}$ and $\sigma_{REML}$. The CRLBs account for 42% to 74% of $\sigma_{REML}$ and for 16% to 50% of $\sigma_{R1,Wc}^S$.
In this work, an estimate of measurement precision for the given setup of in vivo metabolite concentrations for the repeatability and the reproducibility was obtained by a REML analysis and a Bland-Altman analysis. These results were then compared with the commonly used CRLBs for the SPECIAL sequence at 7 T. It was shown that CRLBs depict only a fraction of the measurement precision, whereas the full measurement precision can be obtained by repeated measurements and statistical modeling. Furthermore, the impact of three adiabatic inversion pulses within the SPECIAL sequence, namely, the conventionally applied HS pulse and two gradient-modulated pulses, GOIA, and WURST, on the repeatability and reproducibility were assessed.

The gradient-modulated pulses require a substantially lower peak RF amplitude to fulfill the adiabatic condition than an HS pulse of the same duration, FWHM...
of the inversion profile, and total pulse power.\textsuperscript{28} This advantage of gradient-modulated pulses can be exploited specifically in applications in which the peak RF power or specific absorption rate are limiting factors. Nevertheless, the higher sensitivity to $\Delta B_0$ and gradient imperfections,\textsuperscript{57} as well as the gradient strength and slew rate limitations of the used system, need to be considered during the planning of an application study using gradient-modulated pulses, to avoid the nominal voxel size not being achieved.

The measured in vivo concentrations of the quantified metabolites are well in line with literature values from the same region\textsuperscript{49} and are very similar for all three SPECIAL variants for most metabolites. However, the concentrations of Glu and tCr obtained with both gradient-modulated SPECIAL versions were significantly higher compared with the obtained concentrations from the HS-SPECIAL measurement, and CRLBs were significantly lower for Asp, NAA, and tCr when measured with GOIA-SPECIAL or WURST-SPECIAL compared to HS-SPECIAL. In phantom measurements, however, concentrations, CRLBs, and CVs were approximately the same for all three SPECIAL versions. Differences in the in vivo measurements were tentatively assigned to the substantially reduced CSD and sharper profile achieved by the gradient-modulated pulses compared to the HS pulse in combination with spatial variations of tissue distributions in the brain. The effect on the other metabolites is likely of a similar nature but not identifiable as unambiguous, due to the overlap of several signals in the respective frequency ranges. The smaller intra-subject CVs of metabolite concentrations, as well as the decreased number of metabolite concentrations that were not detected by LCModel, indicate that the reduced CSD and the sharper pulse profile of the inversion in SPECIAL using either a GOIA or a WURST pulse, have a positive effect on the robustness of the LCModel quantification compared to an HS pulse, especially for low-concentration metabolites. It is expected that the same effect of improved fit-robustness could be observed, if the CSD of the excitation and refocusing pulse were also reduced.

Note that due to the small sample size, which is not guaranteed to follow a normal distribution,\textsuperscript{58} the statistical significance was assessed by a non-parametric statistical test, namely, the Wilcoxon signed-rank sum test. This approach provides more conservative estimates than the parametric pendant, the paired t-test.

The assessment of reproducibility in MRS has received increased attention lately, such as the comparability of different scanners and sites or the test–retest reproducibility evaluated with CVs.\textsuperscript{49,59–61} However, the unbalanced nested study design used in this study extends this concept to allow the estimation of realistic SDs of in vivo metabolite concentrations through Bland-Altman analysis and REML analysis for the first time.

| Volunteer | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Voxel overlap/% | 83.2 ± 4.0 | 83.9 ± 4.7 | 83.6 ± 4.2 | 82.7 ± 4.5 | 85.4 ± 7.7 | 90.6 ± 1.0 | 88.8 ± 4.9 | 81.4 ± 9.4 | 87.4 ± 3.4 |
| CSF fraction/% | 1.9 ± 0.7 | 1.7 ± 0.1 | 2.2 ± 0.5 | 1.3 ± 0.2 | 1.9 ± 0.7 | 2.2 ± 0.3 | 0.4 ± 0.1 | 7.1 ± 0.4 | 2.3 ± 1.0 |
| GM fraction/% | 78.0 ± 1.2 | 74.8 ± 2.2 | 77.7 ± 0.9 | 75.4 ± 1.3 | 69.2 ± 1.0 | 75.5 ± 0.6 | 78.3 ± 0.9 | 72.5 ± 3.8 | 78.1 ± 0.5 |
| WM fraction/% | 20.1 ± 1.8 | 23.4 ± 2.2 | 20.1 ± 0.7 | 23.4 ± 1.2 | 28.8 ± 0.6 | 22.2 ± 0.7 | 21.2 ± 1.0 | 19.2 ± 3.4 | 19.6 ± 0.8 |

FIGURE 3: Voxel overlap in the sagittal view for subject 6 (highest overlap). The color map indicates the number of overlapping voxels (i.e., all four voxels overlap in the yellow region).
Although the metabolite concentrations obtained after measurement, post-processing, and absolute quantification are certainly the clinically relevant results and hence, the precision of these values are important for clinical assessments, here the Bland-Altman analysis of the spectral shape was introduced to obtain a complementary measure for reproducibility of in vivo MRS. Since all spectral differences contribute to the final data point in the Bland-Altman plot, and it is not influenced by inaccuracies that might be inherent to the used quantification model, this analysis provides additional information on the reproducibility independent of the quantification pipeline. Although the visually assessed spectral quality, the water linewidth, and its SNR were similar for all three SPECIAL variants (indicating a similar performance of all investigated pulses), the Bland-Altman analysis of the spectral shape reveals subtle differences among the different SPECIAL variants in vivo. Thus, it was demonstrated that the use of the GOIA and WURST pulses for adiabatic inversion in SPECIAL did result in higher repeatability of
the spectral shape. However, inaccuracies in VOI positioning, potential differences in calibration, and other effects appear to outweigh this benefit. The observed increase of $\sigma_{R1,We}^2$ compared with $\sigma_{R1,Mc}^2$ might have several reasons: (1) some volunteers being scanned by different operators in the first and second session, which might have influenced the combined reproducibility of voxel positioning; (2) day-to-day variation of the scanner performance; and (3) intra-subject physiological changes between the two sessions. It is usually aimed to minimize the effects caused by the first two points. Especially effects on the overall variance of the metabolite concentration, caused by differences in voxel positioning, might be mitigated by the use of automated voxel positioning routines, as described by Dou et al.\textsuperscript{62} In this study, it was decided to perform a manual voxel positioning with reference to anatomical landmarks, to reflect the workflow in many clinical studies.\textsuperscript{63} Although operators within this study carefully aimed to position the voxel
as reproducibly as possible, a mean voxel overlap of only 85.2\% could be achieved, which is lower than was demonstrated to be feasible with automated voxel positioning routines\(^\text{62}\) and which is expected to negatively affect the precision obtained in this study. Nevertheless, this work outlines a framework that allows quantifying the influence of measures taken to reduce the measurement precision, such as the mentioned automated voxel position routines. Effects caused by actual physiological changes, on the other hand, may represent the answer to the clinical research question at hand, such as in longitudinal studies. Furthermore, as MRS progresses toward broader clinical use, the need to determine 'normal' ranges of these physiological variations arises, as well as deviations thereof, to be of use as a diagnostic tool on a single-subject basis.

It should be noted that poor B\(_0\) shimming would lead to a larger linewidth of the metabolites, which would hamper correct quantification, as the overlap of adjacent peaks would be increased. This would result in larger CRLBs, and likely in larger SD of the calculated metabolite concentrations. However, if poor B\(_0\) shimming is reproducible, the Bland-Altman analysis of the spectral shape would not be expected to change substantially.

It is worth noting when looking at the Bland-Altman analysis of repeatability and reproducibility of the calculated metabolite concentrations, that while \(\sigma_{\text{REML}}^{\text{RA}} < \sigma_{\text{REML}}^{\text{BA}}\) the effect is not as pronounced as found in the analysis of the spectral shape, and that no consistent trend can be observed between \(\sigma_{\text{REML}}^{\text{RA}}\) and \(\sigma_{\text{REML}}^{\text{BA}}\). This indicates that inaccuracies in the used fitting model and differences in
fit quality ‘mask’ the differences between the two investigated reproducibility scenarios.

A similar effect can be observed in the results from the REML analysis. The REML fit prevents negative estimates, as a negative variance would not make physical sense. This, in combination with the modest degrees of freedom and relatively large within-group effects, leads to several variance contributions of both reproducibility scenarios to be estimated as nominally zero— or very close to. Note that either $\sigma_{R1,M}^{\text{REML}}$ or $\sigma_{R1,W}^{\text{REML}}$ of the different metabolites is estimated as zero, but never both at the same time. This does not mean that the group means from both scenarios—either $R_0$ and $R_{1,Mc}$ if $\sigma_{R1,M}^{\text{REML}} = 0$, or $R_{1,Mc}$ and $R_{1,Wc}$ if $\sigma_{R1,W}^{\text{REML}} = 0$—are identical; it only reflects that they are closer together than expected from the within-group variance. Hence, for most of the metabolites, the two examined reproducibility scenarios and the effect of their respective variance contribution on the measurement precision cannot be clearly disentangled—probably due to the small sample size of only 9 volunteers.

Although the results from the REML analysis exhibit a small tendency toward lower SDs in data obtained with gradient-modulated inversion pulses, these differences are neither statistically significant nor generally consistent. Hence, to strengthen the investigation on the MDCs for different metabolites, the SDs for the repeatability $\sigma_{R0}^{\text{REML}}$ and the combined reproducibility $\sigma_{R1,Wc}^{\text{REML}}$ of the three investigated pulses were pooled. The variances derived as $\sigma_{R1,Wc}^{\text{REML}}$ then allowed for the calculation of the MDC for the given setup.

As the REML analysis contains a multi-parameter fit model with multiple contributions to the total variance, it properly accounts for the unbalanced nested study design

| Metabolite | MDC/µmol g$^{-1}$ |
|------------|-------------------|
| Asp        | 1.87              |
| GABA       | 1.20              |
| Gln        | 1.22              |
| Glu        | 2.23              |
| GSH        | 0.66              |
| Ins        | 1.66              |
| Lac        | 0.59              |
| NAA        | 2.16              |
| NAAG       | 0.92              |
| PE         | 0.40              |
| tCho       | 1.92              |
| tCr        | 1.46              |
| Tau        | 0.85              |

and weighs incomplete data sets without discarding the information completely. Therefore, the SDs obtained by REML analysis are considered to be a more reliable precision estimate than the SDs obtained by Bland-Altman analysis. Nevertheless, the Bland-Altman analysis provides a valuable consistency check with similar trends observed, aiding in the interpretation of the rather complex REML analysis results. Furthermore, these differences are expected to decrease for a larger sample size. Comparison of these results with the CRLBs reveals that CRLBs only account for a fraction of the measurement variance. This is not surprising, however, as CRLBs are the “lowest possible standard deviations of all unbiased model parameter estimates obtained from the data,”12 which are limited to the variance contributions resulting from the noise level, the overlap of different peaks, and the metabolite fitting model64; they do not reflect the SD of data that would be obtained by repeated measurements.31 Hence, it should be noted that the framework presented here is not aiming to replace the use of CRLBs—as repeated measurements to obtain SDs of metabolite concentrations will remain impossible in most clinical settings—but provides complementary information for a better understanding of the precision of metabolite concentrations obtained by MRS. Nevertheless, strong correlations are found among all three measures for the different metabolites. This finding strengthens and underpins approaches using the CRLBs as weights in statistical analysis, as suggested by Miller et al.55

Although the obtained SDs and MDCs are rather large, the correlation between the two repeatability acquisitions across the subjects (e.g., Supporting Information Figure S16) demonstrates the general capability of the used $^1$H MRS method to quantify concentration differences between individual subjects.

The chosen SPECIAL localization technique required an add-subtract scheme to obtain full spatial localization. While this allows for very short TEs and at the same time retains the maximum attainable echo amplitude at a given TE, this renders the obtained spectra more susceptible to motion artifacts66 than, for example, sLASER, which achieves full localization for each transient. This may negatively impact the precision that can be achieved compared to single-shot localization techniques, especially when looking at patient cohorts instead of healthy volunteers. Furthermore, there is an ongoing debate in the MRS community regarding the tradeoff between reduced CSD achievable with adiabatic refocusing and minimal TE, and which one is favorable. On the one hand, some experts argue that the reduced CSD of sSPECIAL66 is expected to result in increased reliability of obtained metabolite concentrations, despite the longer TE.67 On the other hand, it was shown that short TEs are especially beneficial for the
reliability of the determination of J-coupled metabolites.\textsuperscript{14}
While the current study did not set out to answer which one of these influences is bigger on the precision, it provides a framework that will allow future studies to compare the precision of metabolite concentrations obtained with different CSDs, TEs, and numbers of transients required for full localization.

The validity of the values derived in this study is certainly limited to the specific setup and methodology used here. The numbers will likely be different for other brain regions, MR scanners, MRS sequences, sequence parameters, $B_0$ and $B_1^+$ calibrations, post-processing pipelines, and fitting models. Nonetheless, this work presents a generally applicable framework to distinguish different contributions to the total measurement variance and to investigate the efficacy of specific measures aiming to reduce individual variance contributions systematically. Finally, this provides the groundwork for a broader implementation of MRS into clinical applications, as inter-subject differences in comparisons of cohorts or intra-subject differences in longitudinal studies can only be reliably distinguished from statistical fluctuations if they are larger than the MDC.

5 | CONCLUSIONS

This work presents a methodology to estimate the measurement precision of in vivo metabolite concentrations obtained by MRS, and consequently the MDCs for 13 metabolite concentrations in vivo for the used setup. Furthermore, a study design and statistical framework are introduced to disentangle different components of the measurement precision that are easily transferrable to a different setup and sequence parameters. This allows us to systematically investigate the efficacy of measures undertaken to reduce the measurement precision, as was demonstrated for the use case of three different inversion pulses.

ACKNOWLEDGMENT
The authors thank Semiha Aydin for assisting during the measurements. This project (18HLT09 –NeuroMET2) has received funding from the EMPIR program co-financed by the Participating States and from the European Union’s Horizon 2020 research and innovation program. LTR was supported by grant number IT7/8-1 of the DFG.

DATA AVAILABILITY STATEMENT
Corresponding R Code for the REML analysis and Python code to generate the Bland-Altman plots of the spectral shape can be found under https://gitlab1.ptb.de/LRiemann/repeatability_reproducibility.git. The raw data, the used .npy files for the Bland-Altman plots, the segmentation results, and the metabolite concentrations used for the REML analysis can be found under 10.5281/zenodo.5500320.

ORCID

\textbf{Layla Tabea Riemann} \url{https://orcid.org/0000-0001-5411-7288}
\textbf{Christoph Stefan Aigner} \url{https://orcid.org/0000-0003-3618-9610}
\textbf{Stephen L. R. Ellison} \url{https://orcid.org/0000-0002-3008-6656}
\textbf{Rüdiger Briühl} \url{https://orcid.org/0000-0003-0111-5996}
\textbf{Ralf Mekle} \url{https://orcid.org/0000-0001-6821-7722}
\textbf{Sebastian Schmitter} \url{https://orcid.org/0000-0003-4410-6790}
\textbf{Oliver Speck} \url{https://orcid.org/0000-0002-6019-5597}
\textbf{Georg Rose} \url{https://orcid.org/0000-0002-2215-150X}
\textbf{Bernd Itermann} \url{https://orcid.org/0000-0002-4087-471X}
\textbf{Ariane Fillmer} \url{https://orcid.org/0000-0003-3333-6603}

REFERENCES

1. Henning A. \textit{In Vivo 1H MRS Applications}. 3rd Edition. Elsevier; 2016.
2. Hundshammer C, Braeuer M, Müller CA, et al. Simultaneous characterization of tumor cellularity and the Warburg effect with PET, MRI and hyperpolarized 13C-MRSI. \textit{Theranostics}. 2018;8:4765-4780.
3. Bok R, Lee J, Sriram R, et al. The role of lactate metabolism in prostate cancer progression and metastases revealed by dual-agent hyperpolarized 13C-MRSI. \textit{Cancers (Basel)}. 2019;11:257.
4. Soher BJ, Doraisswamy PM, Charles HC. A review of 1H MR spectroscopy findings in Alzheimer’s disease. \textit{Neuroimaging Clin N Am}. 2005;15:847-852.
5. Emir UE, Tuite PJ, Öz G. Elevated pontine and putamenal gaba levels in mild-moderate Parkinson disease detected by 7 Tesla proton MRS. \textit{PLoS One}. 2012;7:e30918.
6. Marsman A, Mandl RCW, Klomp DWJ, et al. GABA and glutamate in schizophrenia: a 7 T 1H-MRS study. \textit{Neuroimage Clin}. 2014;6:398-407.
7. Puts NAJ, Edden RAE. In vivo magnetic resonance spectroscopy of GABA: a methodological review. \textit{Prog Nucl Magn Reson Spectrosc}. 2012;60:29-41.
8. Tedeschi G, Bertolino A, Righini A, et al. Brain regional distribution pattern of metabolite signal intensities in young adults by proton magnetic resonance spectroscopic imaging. \textit{Neurology}. 1995;45:1384-1391.
9. Henning A. Proton and multinuclear magnetic resonance spectroscopy in the human brain at ultra-high field strength: a review. \textit{Neuroimage}. 2018;168:181-198.
10. Kantarci K, Weigand SD, Petersen RC, et al. Longitudinal 1H MRS changes in mild cognitive impairment and Alzheimer’s disease. \textit{Neurobiol Aging}. 2007;28:1330-1339.
11. Kumar J, Liddle EB, Fernandes CC, et al. Glutathione and glutamate in schizophrenia: a 7T MRS study. \textit{Mol Psychiatry}. 2020;25:873-882.
12. Cavassila S, Deval S, Huegen C, Van Ormondt D, Graveron-Demilly D. Cramér-Rao bounds: an evaluation tool for quantification. *NMR Biomed.* 2001;14:278-283.

13. Cavassila S, Deval S, Huegen C, Van Ormondt D, Graveron-Demilly D. Cramér-Rao bound expressions for parametric estimation of overlapping peaks: influence of prior knowledge. *J Magn Reson.* 2000;143:311-320.

14. Landheer K, Juchem C. Optimization of echo time choice for seven common MRS pulse sequences through minimization of expected Cramér-Rao lower bounds. In: Proceedings of the 28th Annual Meeting of ISMRM [Online], 2020. Abstract #369.

15. Wilson M, Andronesi O, Barker PB, et al. Methodological consensus on clinical proton MRS of the brain: review and recommendations. *Magn Reson Med.* 2019;82:527-570.

16. Zhong K, Ernst T. Localized in vivo human 1H MRS at very short echo times. *Magn Reson Med.* 2004;52:898-901.

17. Mekle R, Mlynárik V, Gambarota G, Hergt M, Krueger G, Gruetter R. MR spectroscopy of the human brain with enhanced signal intensity at ultrashort echo times on a clinical platform at 3T and 7T. *Magn Reson Med.* 2009;61:1279-1285.

18. Mlynárik V, Gambarota G, Frenkel H, Gruetter R. Localized short-echo-time proton MR spectroscopy with full signal-intensity acquisition. *Magn Reson Med.* 2006;56:965-970.

19. Tannus A, Garwood M. Adiabatic pulses. *NMR Biomed.* 1997;10:423-434.

20. Bernstein. *Handbook of MRI Pulse Sequences.* Elsevier; 2004.

21. Demilly D. Cramér-Rao bounds: an evaluation tool for quantification. *J Econ Entomol.* 1946;39:269.

22. Dehghani M, Edden R, Near J. Simultaneous ultra-short TE-MRS in two voxels using a SPECIAL sequence with Hadamard encoding. In: Proceedings of the 28th Annual Meeting of ISMRM [Online], 2020. Abstract #491.

23. Dhamala E, Abdelkafi I, Nguyen M, Hennessy TJ, Nadeau H, Near J. Validation of in vivo MRS measures of metabolite concentrations in the human brain. *NMR Biomed.* 2019;32:1-15.

24. Xin L, Schaller B, Mlynarik V, Lu H, Gruetter R. Proton T1 relaxation times of metabolites in human occipital white and gray matter at 7 T. *Magn Reson Med.* 2013;69:931-936.

25. Heo H, Kim S, Lee HH, et al. On the utility of short echo time (TE) single voxel 1H-MRS in non-invasive detection of 2-hydroxyglutamate (2HG): challenges and potential improvement illustrated with animal models using MURUI and LCModel. *PLoS One.* 2016;11:1-18.

26. Gambarota G, Mekle R, Xin L, et al. In vivo measurement of glycine with short echo-time 1H MRS in human brain at 7 T. *Magn Reson Mater Phys Biol Med.* 2009;22:1-4.

27. Tannus A, Garwood M. Improved performance of frequency-swept pulses using offset-independent adiabaticity. *J Magn Reson A.* 1996;120:133-137.

28. Andronesi OC, Ramadan S, Ratai EM, Jennings D, Mountford CE, Sorensen AG. Spectroscopic imaging with improved gradient modulated constant adiabaticity pulses on high-field clinical scanners. *J Magn Reson.* 2011;23:1-7.

29. Deelchand DK, Berrington A, Noeske R, et al. Across-vendor standardization of semi-LASER for single-voxel MRS at 3T. *NMR Biomed.* 2021;34:e4218.

30. Chmelik M, Just Kukurová I, Gruber S, et al. Fully adiabatic 31P 2D-CI with reduced chemical shift displacement error at 7 T—GOIA-1D-ISIS/2D-CSI. *Magn Reson Med.* 2013;69:1233-1244.

31. Kreis R, Boer V, Choi I-Y, et al. Terminology and concepts for the characterization of in vivo MR spectroscopy methods and MR spectra: background and experts' consensus recommendations. *NMR Biomed.* 2021;34:e4347.

32. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther.* 2006;86:735-743.

33. Rand A, Aigner CS, Kunisch K, Stollberger R. Simultaneous multislice refocusing via time optimal control. *Magn Reson Med.* 2018;80:1416-1428.

34. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol.* 2008;31:466-475.

35. Marques JP, Kober T, Krueger G, van der Zwaag W, Van de Moortele PF, Gruetter R. MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *Neuroimage.* 2010;49:1271-1281.

36. Fillmer A, Kirchner T, Cameron D, Henning A. Constrained image-based B0 shimming accounting for “local minimum traps” in the optimization and field inhomogeneities outside the region of interest. *Magn Reson Med.* 2015;73:1370-1380.

37. Nassirpour S, Chang P, Fillmer A, Henning A. A comparison of optimization algorithms for localized in vivo B0 shimming. *Magn Reson Med.* 2018;79:1145-1156.

38. Tkáč I, Starcuk Z, Choi I-Y, Gruetter R. In vivo 1H NMR spectroscopy of rat brain at 1 ms echo time. *Magn Reson Med.* 1999;41:649-656.

39. Mao J, Yang H, Fitzsimmons JR. Slice profile improvement for a clinical MRI system. *Magn Reson Imaging.* 1990;8:767-770.

40. Tkáč I, Gruetter R. Methodology of 1H NMR spectroscopy of the human brain at very high magnetic fields. *Appl Magn Reson.* 2005;29:139-157.

41. Brown MA. Time-domain combination of MR spectroscopy data acquired using phased-array coils. *Magn Reson Med.* 2004;52:1207-1213.

42. Provencher SW. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magn Reson Med.* 1993;30:672-679.

43. Soher BJ, Semanchuk P, Todd D, Steinberg J, Young K. VeSPA: integrated applications for RF pulse design, spectral simulation and MRS data analysis. *J Magn Reson Ser A.* 2011;12:75-105.

44. Cudalbu C, Behar KL, Bhattacharyya PK, et al. Contribution of macromolecules to brain 1 H MR spectra: experts’ consensus recommendations. *NMR Biomed.* 2021;34:e4393.

45. Friston K, Ashburner J, Kiebel S, Nichols T, Penny W, eds. *Statistical Parametric Mapping.* 1st Edition. Elsevier; 2007.

46. Van Rossum G. *Python Tutorial*; 1995.

47. Li Y. T1 and T2 metabolite relaxation times in normal brain at 3T and 7T. *J Mol Imaging Dyn.* 2013;02:1-5.

48. Wyss M, Kirchner T, Ringenbach A, Prüssmann K, Henning A. Relaxation parameter mapping adapted for 7T and validation against optimized single voxel MRS. In: Proceedings of the 26th Annual Meeting of ISMRM, Salt Lake City, Utah, 2013. pp 47-48.

49. Terpstra M, Cheong I, Lyu T, et al. Test-retest reproducibility of neurochemical profiles with short-echo, single-voxel MR spectroscopy at 3T and 7T. *Magn Reson Med.* 2016;76:1083-1091.

50. Wilcoxon F. Individual comparisons of grouped data by ranking methods. *J Econ Entomol.* 1946;39:269.
51. Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327:307-310.

52. Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. *J Am Stat Assoc*. 1977;72:320-338.

53. R Core Team. *R: A Language and Environment for Statistical Computing*. 2020.

54. Pinheiro J, Bates D, Core R, *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-147. https://cran.r-project.org/package=nlme. Accessed September 01, 2021.

55. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek ALM. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res*. 2001;10:571-578.

56. ISO. *Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement*. ISO. 2008. http://www.bipm.org/en/publications/guides/gum.html. Accessed September 01, 2021.

57. Aigner CS, Rund A, Abo Seada S, et al. Time optimal control-based RF pulse design under gradient imperfections. *Magn Reson Med*. 2020;83:561-574.

58. Campbell MJ, Swinscow TDV. Statistics at square one. *Br Med J*. 1976;1:1240.

59. Deelchand DK, Adanyeguh IM, Emir UE, et al. Two-site reproducibility of cerebellar and brainstem neurochemical profiles with short-echo, single-voxel MRS at 3T. *Magn Reson Med*. 2015;73:1718-1725.

60. Wijnen JP, Rowland LM, Oeltzschner G, Barker PB, Workman C1, Smith GS. Reproducibility of brain MRS in older healthy adults at 7T. *NMR Biomed*. 2019;32:1-8.

61. 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.

62. Dou W, Speck O, Benner T, et al. Automatic voxel positioning for MRS at 7 T. *Magn Reson Mater Phys Biol Med*. 2015;28:259-270.

63. Bai X, D. Harris A, Gong T, et al. Voxel placement precision for GABA-edited magnetic resonance spectroscopy. *Open J Radiol*. 2017;7:35-44.

64. Near J, Harris AD, Juchem C, et al. Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations. *NMR Biomed*. 2021;34:e4257.

65. Miller JJ, Cochlin L, Clarke K, Tyler DJ. Weighted averaging in spectroscopic studies improves statistical power. *Magn Reson Med*. 2017;78:2082-2094.

66. Fuchs A, Luttje M, Boesiger P, Henning A. SPECIAL semilASER with lipid artifact compensation for 1H MRS at 7 T. *Magn Reson Med*. 2013;69:603-612.

67. Öz G, Deelchand DK, Wijnen JP, et al. Advanced single voxel 1H magnetic resonance spectroscopy techniques in humans: experts’ consensus recommendations. *NMR Biomed*. 2020;e4236.

**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher’s website.

**FIGURE S1** Comparison of (A) obtained metabolite concentrations, (B) CRLBs, and (C) CVs for all metabolites in the phantom using the SPECIAL sequence with either HS (blue), GOIA (orange), or WURST (green) as adiabatic inversion pulses. The error bars indicate the SD

**FIGURE S2-S15** Bland-Altman plots for the concentration differences over the three pulses for the $R_{BA}$, the $R_{1,M}$, and the $R_{1,W}$ scenario (upper, middle, and lower row, respectively). The blue points indicate the SPECIAL sequence measured with the HS pulse, while the orange ones indicate GOIA and the green ones WURST. Note that for the $R_{1,W}$ scenario, the first measurements of both days were used

**FIGURE S16** Correlation plots for the $R_0$ scenario over all volunteers and RF inversion pulses for (A) tCr, (B) Glu, (C) tCho, and (D) NAA. The red line indicates the fitted curve with the $R^2$ value and the slope, while the gray lines indicate the $\pm$SD of the fitted curve

**TABLE S1** Water linewidth, necessary transmitter voltage, and SNR$_{\text{water}}$ for the three pulses

**TABLE S2** Values derived from the REML analysis for each inversion pulse type and each metabolite: The SD of the reproducibility scenarios $\sigma_{R_{1,W}}^{\text{REML}}$ and $\sigma_{R_{1,M}}^{\text{REML}}$, the combined reproducibility SD $\sigma_{R_{1,W,c}}^{\text{REML}}$, as well as the SD of the repeatability scenario $\sigma_{R_0}^{\text{REML}}$ are given. N denotes the number of observations included in the mixed effect analysis for each group. To compare the values derived with the REML analysis, the CRLBs, as well as $\sigma_{R_{1,W,c}}^{\text{BA}}$ are given

**How to cite this article:** Riemann LT, Aigner CS, Ellison SLR, et al. Assessment of measurement precision in single-voxel spectroscopy at 7 T: Toward minimal detectable changes of metabolite concentrations in the human brain in vivo. *Magn Reson Med*. 2022;87:1119–1135. doi:10.1002/mrm.29034