



Published in final edited form as:

*NMR Biomed.* 2017 July ; 30(7): . doi:10.1002/nbm.3710.

## When Are Metabolic Ratios Superior to Absolute Quantification? A Statistical Analysis

Sarah E. Hoch<sup>1</sup>, Ivan I. Kirov<sup>2</sup>, and Assaf Tal<sup>3</sup>

<sup>1</sup>Department of Diagnostic Imaging, Sheba Medical Center, Ramat Gan, 5262100, Israel

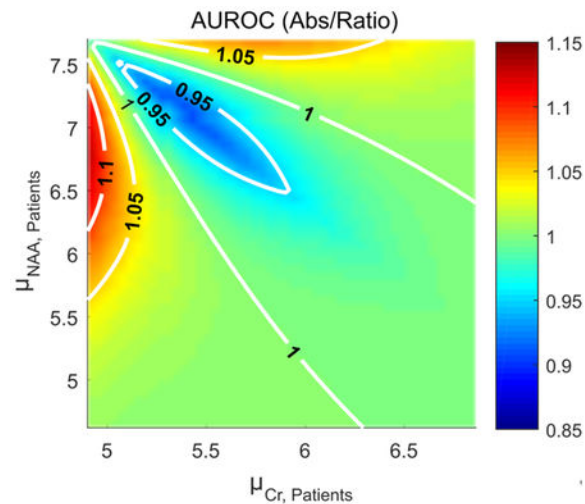
<sup>2</sup>Bernard and Irene Schwartz Center for Biomedical Imaging and Center for Advanced Imaging Innovation and Research (CAI<sup>2</sup>R), Department of Radiology, New York University School of Medicine, 660 1<sup>st</sup> Avenue, New York, NY 10016, USA

<sup>3</sup>Department of Chemical Physics, Weizmann Institute of Science, 234 Herzl St., Rehovot 7610001, Israel

### Abstract

Metabolite levels measured using magnetic resonance spectroscopy (MRS) are often expressed as ratios rather than as absolute concentrations. However, the inter-subject variability of the denominator metabolite can introduce uncertainty into a metabolite ratio. In a clinical setting, there are no guidelines on whether ratios or absolute quantification should be used for a more accurate classification of normal versus abnormal results based on their statistical properties. In a research setting, choosing one over the other can have significant implications on sample size, which must be factored in at the study design stage. Herein, we derive the probability distribution function for the ratio of two normally distributed random variables, and present analytical expressions for comparing ratios to absolute quantification in terms of both sample size and area under the receiver operator characteristic curve. The two approaches are compared for typical metabolite values found in the literature, and their respective merits are illustrated using previously acquired clinical MRS data in two pathologies: mild traumatic brain injury and multiple sclerosis. Our analysis shows that deciding between ratios and absolute quantification is non-trivial: in some cases, ratios might offer a reduction in sample size, while in others, absolute quantification might prove more desirable for individual (i.e. clinical) use. The decision is straightforward and exact guidelines are provided in the text, given that population means and standard deviations of numerator and denominator can be reliably estimated.

### Graphical Abstract



Metabolite ratios can either improve or diminish the usefulness of MRS, compared to absolute quantification, depending on the change in the ratio's numerator and denominator. An analysis of when to choose one over the other is presented, based on considerations of sample size requirements and area under the receiver operator characteristic (AUROC).

### Keywords

Magnetic resonance spectroscopy; MRS; quantification; metabolite ratios; absolute quantification; receiver operator characteristic; ROC; sample size

### Introduction

The quantification of MRS data requires some form of reference signal. This signal can be external, as in a replacement phantom (1,2) or stable signal injected into the coil's RF system (3,4); It can also be internal, as in the water signal from an unsuppressed acquisition (1,5) or another metabolite present in the spectrum, as seen in clinical applications (6-11). If the concentration of the reference signal is known, a statement can be made about the absolute concentration of the metabolite being quantified; otherwise, only a ratio can be reported. Since ratios do not reveal whether an observed difference is due to a change in the numerator or in the denominator, absolute quantification is deemed preferable whenever possible, despite often requiring additional scans and calibration steps (12,13). This increased complexity sometimes leads practitioners to instead favor metabolic ratios, which also account for hardware imperfections and cerebrospinal fluid content.

Another instance when metabolic ratios are preferred is in cases when metabolites are expected to shift in opposing directions, as reported with proton MRS ( $^1\text{H}$ -MRS) in many neurological pathologies, including multiple sclerosis (MS) in which the n-acetyl-aspartate (NAA) levels decrease while creatine (Cr), myo-inositol (mI) and choline (Cho) increase (14); dementia, in which NAA decreases while Cr increases (15); and brain tumors, in which Cr and NAA often decrease while Cho increases (16). Metabolite ratios capitalize on this trend, since the ratio of two such quantities  $A/B$  will change in pathology  $A/B \rightarrow (A$

+ A)/(B- B) markedly more than either numerator (A→A+ A) or denominator (B→B- B).

As previously noted (17,18), some metabolic ratios display increased coefficients of variation (CVs) compared to absolute quantification due to the combined variances of both numerator and denominator. To adequately decide whether ratios are statistically superior to absolute quantification, however, CVs alone are insufficient and one must examine the expected change and the CVs simultaneously. Despite the prevalence of ratios in the research and clinical literature, no attempt has been made to examine this question in depth. Herein we contrast absolute quantification with metabolite ratios by examining their statistical properties. We begin by deriving a simplified probability distribution function (pdf) of the ratio of two metabolites. This is used to determine two key quantities: the reduction in sample size required to observe a change in a cohort study, and the total area underneath the Receiver Operator Characteristic (AUROC). Minimizing sample size reduces both the cost and complexity of studies which use <sup>1</sup>H-MRS to define clinical end points. The AUROC equals the probability that a test correctly classifies those with and without the disease (19,20); it is often used as indicator of the quality of a classifier. Both metrics are numerically computed and examined for a range of physiologically relevant metabolite values, based on literature reports and on two previously acquired datasets from MS (14) and mild traumatic brain injury (mTBI) (21).

## Theory

### The Probability Distribution of a Metabolic Ratio

The distribution of most metabolites is approximately normal. The probability distribution function  $p(R)$  of the ratio  $R=X_1/X_2$  of two normally distributed random variables,  $X_1 \sim N(\mu_1, \sigma_1^2)$ ,  $X_2 \sim N(\mu_2, \sigma_2^2)$  has been previously derived (22):

$$\begin{aligned}
 p(R) &= g(R)\exp[h(R)] \\
 g(R) &= \frac{b(R)}{\sqrt{2\pi}\sigma_1\sigma_2 a^3(R)} \\
 h(R) &= \frac{b^2(R) - c \cdot a^2(R)}{2(1 - \rho^2)a^2(R)} \\
 b(R) &= \frac{\mu_1}{\sigma_1^2}R - \frac{\rho(\mu_1 + \mu_2 R)}{\sigma_1\sigma_2} + \frac{\mu_2}{\sigma_2^2} \quad [1] \\
 a(R) &= \sqrt{\frac{R^2}{\sigma_1^2} - \frac{2\rho R}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2}} \\
 c &= \frac{\mu_1^2}{\sigma_1^2} - \frac{2\rho\mu_1\mu_2}{\sigma_1\sigma_2} + \frac{\mu_2^2}{\sigma_2^2}
 \end{aligned}$$

where  $-1 \leq \rho \leq 1$  is the correlation coefficient between  $X_1$  and  $X_2$ . Often, clinical data will have CVs  $\leq 0.25$ . Under this assumption, Eq. [1] can be simplified as follows: The function

$h(R)$  in the exponent has two extremal points,  $R_1$  and  $R_2$  (solutions of  $h'(R)=0$ ), at which (Fig. 1a):

$$h\left(R_1 = \frac{\mu_1}{\mu_2}\right) = 0$$

$$h\left(R_2 = -\frac{\sigma_1 c v_1 - \rho \cdot c v_2}{\sigma_2 c v_2 - \rho \cdot c v_1}\right) = \frac{c v_2^2 - 2\rho \cdot c v_1 c v_2 + c v_1^2}{2c v_1^2 c v_2^2 (\rho^2 - 1)} \quad [2]$$

with  $CV_i = \sigma_i/\mu_i$  the coefficient of variation for the numerator ( $i=1$ ) or denominator ( $i=2$ ). Second derivatives indicate  $R_1$  is a maximum while  $R_2$  is a minimum (Fig. 1a). Furthermore, taking the limits, one finds  $h(R) \xrightarrow{R \rightarrow \pm\infty} -\frac{1}{2c v_2^2}$ . Under the assumption  $CV_2 \ll 1$ , this

implies  $\exp(h(R))$  becomes non-negligible only in the vicinity of the maximum  $R=R_1$  (Fig.

1b). We can therefore expand  $h(R)$  in a series around  $R_1$ ,  $h(R) \approx h(R_1) + \frac{1}{2} h''(R_1) \left(R - \frac{\mu_1}{\mu_2}\right)^2$ ,

with negligible higher order terms since  $h(R)$  is heavily damped by the exponent. The slowly-varying function  $g(R)$ , shown in Fig. 1d, can simply be substituted by its value at the extremum,  $g(R) \approx g(R_1)$ . Following normalization,  $\int_{-\infty}^{\infty} p(R) dR = 1$ , a Gaussian distribution is obtained:

$$R \sim N\left(\frac{\mu_1}{\mu_2}, \left(\frac{\mu_1}{\mu_2}\right)^2 \left(c v_1^2 - 2\rho \cdot c v_1 \cdot c v_2 + c v_2^2\right)\right). \quad [3]$$

The approximate distribution (Eq. [3]) is plotted in Fig. 1c against the true distribution,  $p(R)$  (Eq. [1]). The assumptions used to derive Eq. [3], namely  $CV_2 \ll 1$  and normality, will be examined below.

### Metrics For Assessing Patient and Control Distributions

A normally distributed ratio allows one to easily calculate other quantities of statistical interest. Given normal patient and control ratio distributions,  $R_p \sim N(\mu_p, \sigma_p^2)$ ,  $R_c \sim N(\mu_c, \sigma_c^2)$ , the sample size required to observe a one-sided change in the mean,  $\mu = \mu_p - \mu_c$ , is:

$$N = \frac{(z_{1-\alpha} \sigma_p - z_{\beta} \sigma_c)^2}{\Delta\mu^2} \quad [4]$$

where  $\alpha$ ,  $1-\beta$  are the statistical significance and power of the study, respectively, and  $\mu_p$ ,  $\mu_c$ ,  $\sigma_c$  and  $\sigma_p$  can be calculated using Eq. [3] given the means and variances of the numerator and denominator metabolites.  $z_x$  is the inverse of the normal cumulative distribution function. In many studies one sets  $\alpha=0.05$ ,  $\beta=0.2$ , and so  $z_{0.2}=-0.842$ ,  $z_{0.95}=1.645$ .

A second quantity of interest is the area underneath the receiver operator characteristic (AUROC). Receiver operating characteristic (ROC) curves are often used to assess the performance of a classifier which decides whether a given result is true (e.g. pathological) or false (e.g. healthy), by plotting its true positive rate (sensitivity) as a function of its false positive rate (1-specificity) (23). A coin toss is characterized by a straight line, with an area under the curve of AUROC=0.5, while any non-random useful classifier will deviate from a straight line and have AUROC>0.5 (Fig. 2). The AUROC is bounded from above by unity, and a higher value indicates a superior classifier. The AUROC equals its discrimination: the probability that a randomly chosen true positive (pathological) specimen outranks a randomly chosen true negative (control) (19). Given two uncorrelated normal distributions for controls and patients,  $R_c \sim N(\mu_c, \sigma_c^2)$ ,  $R_p \sim N(\mu_p, \sigma_p^2)$ , the AUROC can be shown to equal:

$$\text{AUROC} = \Phi\left(\frac{|\mu_c - \mu_p|}{\sqrt{\sigma_c^2 + \sigma_p^2}}\right) \quad [5]$$

where  $\Phi(z)$  is the cumulative distribution function of the Gaussian distribution  $N(0,1)$ .

## Materials and Methods

### Absolute Quantification vs. Ratios: Numerical Simulations

To determine when metabolite ratios improve sample size requirements and diagnostic value, we computed  $N$  (Eq. [4]) and the AUROC (Eq. [5]), respectively, for NAA, Cr and NAA/Cr. Literature values for white matter (WM) NAA ( $7.7 \pm 0.6$  mM) and Cr ( $4.9 \pm 0.5$  mM) in healthy controls were used (24). We have confined ourselves to the regime in which NAA decreased and Cr increased, for which ratios would normally be deemed statistically beneficial. Three cases were examined, in which patient standard deviations decreased ( $\sigma_{\text{NAA,P}}=0.4$ ,  $\sigma_{\text{Cr,P}}=0.3$ ), remained unchanged ( $\sigma_{\text{NAA,P}}=0.6$ ,  $\sigma_{\text{Cr,P}}=0.5$ ) and increased ( $\sigma_{\text{NAA,P}}=0.9$ ,  $\sigma_{\text{Cr,P}}=0.9$ ) standard deviations in the patient population. The patient WM NAA and Cr values were varied by up to 40% (typical of many pathologies), keeping the inter-subject coefficient of variation smaller than 25% for the denominator, Cr, and hence validating use of Eq. [3]. The sample sizes  $N_{\text{NAA}}$ ,  $N_{\text{Cr}}$  and  $N_{\text{ratio}}$  required to observe the single-sided change between controls and patients were computed using a statistical significance of 5% and power of 80%.

### Absolute Quantification vs. Ratios: In-Vivo Comparisons

Global WM changes in NAA, Cr and NAA/Cr were examined using data from two previously acquired datasets: a longitudinal study of relapsing-remitting MS, and a cross-sectional mTBI study. The cohort and data analysis pipeline are described elsewhere (14,21). To briefly summarize: the MS cohort consists of eighteen patients (5 men, 13 women) aged 21-45, scanned every 6 months for a period of 3 years, and 10 age- and gender-matched healthy controls scanned yearly; the mTBI cohort consists of twenty-six patients with closed-head trauma (5 women, 21 men) aged  $33 \pm 11$  years, scanned up to 55 days from injury, and thirteen age- and gender-matched healthy controls (5 women, 8 men). All scans

were approved by the Institutional Review Board of New York University School of Medicine. Both cohorts were scanned on a 3T whole-body MR imager (Tim-Trio, Siemens AG, Erlangen Germany) using hybrid Hadamard-PRESS 3D  $^1\text{H}$ -MRS covering a 360 cm<sup>3</sup> volume. Linear regression was used to calculate the global WM values of NAA, Cr, Cho and mI (24). Data was processed using in-house software and fit using SITools-FITT (25), and phantom replacement was used for absolute quantification, based on the signal formula

$$s = \Lambda_0 \cdot c \cdot v_{ref} \cdot v_{voxel} \cdot \left(1 - e^{-TR/T_1}\right) \cdot e^{-TE/T_2} \quad [6]$$

where  $\Lambda_0$  is a proportionality constant which depends on the electronics and coils;  $C$  the concentration, in mM;  $V_{voxel}$  the voxel volume, in mL;  $V_{ref}$  the reference voltage, in volts; and  $T_1, T_2$  the metabolite-specific relaxation constants in question, in either the phantom or the in-vivo tissue.  $B_{1+}$  and  $B_{1-}$  variations in the 360 mL spectroscopic volume were not accounted for. Note that Eq. [6] implies taking the ratio of two absolutely quantified metabolites is completely equivalent to the ratio of the original metabolic signals following  $T_1, T_2$  correction, since the phantom contributions cancel out among both metabolites.

A Shapiro-Wilks test with a significance level of  $\alpha=0.05$  was used to validate the normality of numerator, denominator and ratio. A two-sided t-test at  $\alpha=0.05$  was used to verify whether one can assume  $\rho = 0$  in Eq. [3]. Estimated population means and standard deviations for NAA and Cr were used to compute required sample sizes and AUROC using Eqs. [4], [5] for NAA, Cr and NAA/Cr (in MS) and NAA, mI and NAA/mI (in mTBI).

## Results

### Absolute Quantification vs. Ratios: Numerical Simulations

Fig. 3 compares both sample size requirements and the AUROC between absolute quantification and metabolite ratios, for all three cases: decreased, unchanged and increased population standard deviations. Plotted are two quantities in the  $(\mu_{NAA}^{(patient)}, \mu_{Cr}^{(patient)})$  plane: the ratio  $\min(N_{NAA}, N_{Cr})/N_{ratio}$  (Figs. 3a,c,e) and  $\max(\text{AUROC}_{NAA}, \text{AUROC}_{Cr})/\text{AUROC}_{ratio}$  (Figs. 3b,d,f), with several contours. A ratio of unity (dashed, red) corresponds to the case in which ratios and absolute quantification perform equally. Ratios are preferable in the strip within the unity contour, while absolute quantification is preferable outside the strip. The plots reveal several general behaviors of ratios.

First, a non-trivial relationship exists between numerator, denominator and ratio distributions, and one must carefully consider both changes in mean values and standard deviations before opting for one or the other. For example, for the case of unchanged standard deviations, when NAA declines by 5% and Cr rises by 5%,  $\lceil N_{Cr} \rceil = 27$ ,  $\lceil N_{NAA} \rceil = 15$  and  $\lceil N_{ratio} \rceil = 11$ ,  $\text{AUROC}_{Cr} = 0.63$ ,  $\text{AUROC}_{NAA} = 0.68$ ,  $\text{AUROC}_{ratio} = 0.71$  (here  $\lceil x \rceil$  denotes the smallest possible integer larger than or equal to  $x$ ). While ratios are superior to absolute quantification in both cases, they clearly have a much more pronounced advantage when considering sample sizes. An opposite situation arises when NAA declines by 15% and Cr rises by 15%, where  $\lceil N_{Cr} \rceil = 3$ ,  $\lceil N_{NAA} \rceil = 2$  and  $\lceil N_{ratio} \rceil = 2$ ,

$AUROC_{Cr}=0.85$ ,  $AUROC_{NAA}=0.91$ ,  $AUROC_{ratio}=0.95$ ; here, sample size requirements are equal but diagnostically, the ratio NAA/Cr still presents an improvement over both NAA and Cr. These examples illustrate a second point: as shown by all cases, the relative merit of ratios vs. absolute quantification depend on the metric considered. The regions within which ratios are preferable to absolute quantification are similar but not equivalent.

Third, ratios become more favorable as the standard deviation in the patient population decreases relative to controls. This is a direct consequence of the decreased variability introduced into the ratio NAA/Cr by dividing by Cr: as the denominator's variability decreases, ratios become more statistically appealing. This same behavior can also be seen by, e.g., considering case II with NAA declining by 10% and Cr remaining unchanged: here,  $\lceil N_{Cr} \rceil = \infty$ ,  $\lceil N_{NAA} \rceil = 4$  and  $\lceil N_{ratio} \rceil = 10$ ,  $AUROC_{Cr}=0.5$ ,  $AUROC_{NAA}=0.82$ ,  $AUROC_{ratio}=0.72$ . NAA/Cr displays decreased performance compared to NAA due to division by the Cr value, which does not change on average but has a non-zero standard deviation (owing to inter-subject variability).

A two-sided sample size calculation would simply scale these numbers, but not change the overall form of Fig. 3.

### Absolute Quantification vs. Ratios: In-Vivo Comparisons

All metabolite distributions examined were normally distributed as revealed by the Shapiro-Wilks test for normality. For the ratios considered, none of the metabolite pairs exhibited statistically significant correlation, leading us to assume  $\rho=0$ . Statistically significant ( $\alpha=0.05$ ) reduction in WM NAA and increase in WM Cho, Cr and mI concentrations were observed in several - but not all - time points in MS patients compared with the pooled mean of controls, in accordance with previously reported results (14) (Table 1). In contrast, the ratio NAA/Cr differs significantly at all time points between patients and controls (Fig. 4); all time points remain significant even as the significance level is lowered to a strict  $\alpha=0.01$  level. Despite its more pronounced statistical significance, NAA/Cr actually underperforms compared to absolute quantification in three of the seven time points (4, 5 and 6), at which the sole Cr measurement offers marginally better sample sizes and AUROCs. This clearly illustrates that statistical significance does not imply statistical superiority in terms of sample size or diagnostic quality, although differences in question might be small.

For the mTBI cohort, only WM NAA was statistically different between patients and controls ( $\alpha=0.05$ ) (Fig. 5). Sample size and AUROC calculations based on the estimated population variance and mean yielded the smallest sample size ( $\lceil N_{NAA} \rceil = 10$ ) and largest  $AUROC=0.7$  for NAA, and slightly inferior results for the ratio NAA/mI ( $\lceil N_{mI} \rceil = 14$ ,  $AUROC=0.68$ ). This clearly illustrates a second point: even though NAA and mI change in opposing directions, their ratio is not preferable to absolute quantification due to the additional variance introduced by dividing by mI.

## Discussion

Normally distributed random variables and their ratios are found throughout science, and their ratios appear in fields as diverse as analog circuit design (26), telecommunications (27),

industrial process monitoring (28) and econometrics (29). The statistical literature contains exact derivations (30-34), as well as various application-driven approximations to these ratios, probability distribution functions, moments and confidence intervals (35-37), for both the correlated and uncorrelated case. Herein, we have derived one particular normal approximation valid for typical, normally-distributed MRS data (Eq. [3]), where metabolite CVs rarely exceed 20-25%. This approximation lends itself to comparing absolute quantification to metabolite ratios in terms of their relevant statistical properties.

It is often implicitly assumed that whenever two metabolites' means shift in an opposing manner, their ratio must improve detection. Herein we have shown that, depending on the means and variances of both numerator and denominator, metabolic ratios can either enhance or diminish the AUROC and sample size requirements in a non-trivial manner. On the one hand, ratios of metabolites with opposite behavior exhibit a greater change to the mean; on the other hand, any inter-subject variability present in the reference metabolite carries over to the ratio. Our conclusions can be considered an extension of previous studies, which have already noted the increased CVs exhibited by ratios, albeit without examining their effect on either sample size or contribution to clinical diagnosis (17).

### When Are Ratios Preferable?

Choosing ratios over absolute quantification – from a statistical point of view – should be weighted carefully, especially since in many pathologies MRS peak areas change only modestly and it is not a-priori clear which metric is better; put differently, the change in the metabolite ratio's mean is on the order of the variability inherent in the denominator. We have provided the reader with concrete tools for assessing precisely which distribution is favorable, depending on whether study design (Eq. [4]) or clinical diagnosis (Eq. [5]) are concerned.

We also note that there are multiple other considerations that affect one's choice: absolute quantification often requires additional scan time and adds complexity, but allows for a clearer physiological interpretation of the underlying changes (albeit not always a statistically superior one). However, whether one should even consider adding this complexity should ultimately rest upon the measurement's usage. For example, in a clinical setting where diagnosis is of paramount importance, ratios might offer greater AUROCs in some cases, rendering absolute quantification unnecessary. It should be stressed that deciding between the two requires knowledge of the mean and standard deviation of both numerator and denominator – e.g., from a previous study which did use absolute quantification.

### Caveats And Pitfalls

Statistical significance should not be assumed to indicate statistical superiority. This was demonstrated in the MS cohort (Table 1). Even though Cr did not show statistical significance at three time points, it still offered both smaller sample sizes and better AUROCs compared to NAA/Cr which was statistically significant at the same three points. This should not be surprising: sample size and AUROCs are calculated using a population's distribution, while statistical significance depends on the sampling distribution; indeed, by



sufficiently lowering the significance level ( $\alpha$ ) any result can be rendered statistically non-significant, without affecting our sample size and AUROC estimations.

When using Eqs. [4], [5], one must know a-priori the means and standard deviations of both numerator and denominator metabolites, as well as their correlation coefficient, for both patient and control populations. These must be obtained from a previous MRS study which has employed absolute quantification for the same pathology. It is important to select reliable studies which employ high quality, robust methodology, in which mean and standard deviation estimates are dominated by inter-subject variation, with a minimal intra-subject component related to measurement error. This can be done by focusing on studies with high intraclass correlation coefficient ( $ICC \approx 1$ ), which quantifies the agreement of measurements on the same subject (38).

Our derivations have assumed both numerator and denominator are normally distributed inter-subject, which must be tested on a per-case basis. All metabolite distributions used herein from the mTBI and MS cohorts, as well as their matched controls, were verified to be normally distributed using the Shapiro-Wilks test. There are multiple tests for normality, and the interested reader is referred to the literature for a comprehensive overview (39,40).

Absolute quantification itself is obtained by taking the ratio of a metabolite to a reference signal, and the preceding framework can be used to analyze the uncertainty these ratios introduce, such as the random physiological variation inherent to reference unsuppressed water scans or inter-subject variability in tissue water content. For example, when performing absolute quantification of liver fat content, the ratio of hepatic fat content to liver water content is sometimes used, with an assumed water content of 0.711 g water/g normal tissue (41). However, this does not take into account an inter-subject variation of approximately 10% (42). This variation enters into the ratio distribution as described by Eq. [3], contributes to the standard deviation of the reported metabolite values and should be accounted for when performing sample size estimation and other statistical analyses.

In the case two metabolites simultaneously increase or decrease, their ratio will always be statistically unfavorable compared to absolute quantification.

## Conclusions

Metabolite ratios reduce sample size requirements and increase statistical significance compared to absolute quantification in almost all cases the numerator and denominator shift in opposition in pathology, despite increases in CVs reported by other authors (17,18). Nevertheless, care must be taken when the shifts are small in size, in which case the denominator's variance may become dominant, making absolute quantification statistically favorable. A precise determination of whether ratios are favorable can be made on a per-case basis using Eqs. [3] and [4].

## Acknowledgments

**Sponsors, Grants & Funding:** Assaf Tal acknowledges the support of the Monroy-Marks Career Development Fund, the Carolito Stiftung Fund, the Leona M. and Harry B. Helmsley Charitable Trust and the historic generosity

of the Harold Perlman Family. This research was supported by a research grant from the Sylvia Schaefer Alzheimer's Research Fund. Ivan Kirov acknowledges NIH funding grants R01 NS097494 and P41 EB017183.

## Abbreviations:

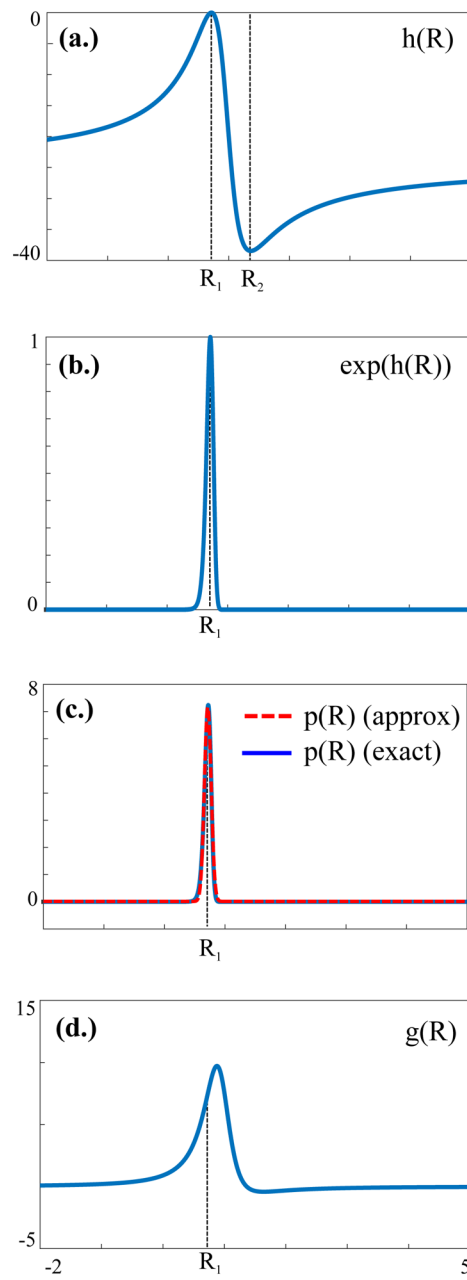
<b>MS</b>	Multiple Sclerosis
<b>mTBI</b>	mild traumatic brain injury
<b>NAA</b>	n-acetyl-aspartate
<b>Cr</b>	creatine, Cho-choline
<b>mI</b>	myo-inositol
<b>CV</b>	coefficient of variation
<b>pdf</b>	probability distribution function
<b>AUROC</b>	area underneath the Receiver Operator Characteristic
<b>ROC</b>	Receiver operating characteristic
<b>WM</b>	white matter

## Bibliography

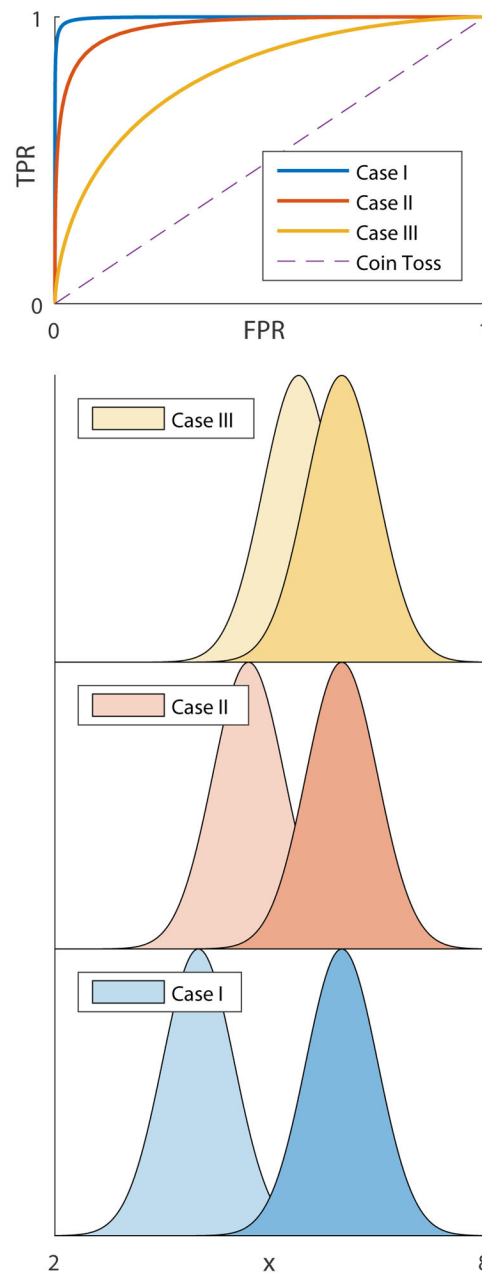
1. Wang X, Salibi N, Fayad LM, Barker PB. Proton magnetic resonance spectroscopy of skeletal muscle: A comparison of two quantitation techniques. *J Magn Reson* 2014;243:81–84. [PubMed: 24792959]
2. Soher BJ, van Zijl PC, Duyn JH, Barker PB. Quantitative proton MR spectroscopic imaging of the human brain. *Magn Reson Med* 1996;35(3):356–363. [PubMed: 8699947]
3. Akoka S, Barantin L, Trierweiler M. Concentration Measurement by Proton NMR Using the ERETIC Method. *Anal Chem* 1999;71(13):2554–2557. [PubMed: 21662801]
4. Heinzer-Schweizer S, De Zanche N, Pavan M, Mens G, Sturzenegger U, Henning A, Boesiger P. In-vivo assessment of tissue metabolite levels using 1H MRS and the Electric REference To access In vivo Concentrations (ERETIC) method. *NMR Biomed* 2010;23(4):406–413. [PubMed: 20101606]
5. Tong Z, Yamaki T, Harada K, Houkin K. In vivo quantification of the metabolites in normal brain and brain tumors by proton MR spectroscopy using water as an internal standard. *Magn Reson Imaging* 2004;22(7):1017–1024. [PubMed: 15288143]
6. Kendall GS, Melbourne A, Johnson S, Price D, Bainbridge A, Gunny R, Huertas-Ceballos A, Cady EB, Ourselin S, Marlow N, Robertson NJ. White matter NAA/Cho and Cho/Cr ratios at MR spectroscopy are predictive of motor outcome in preterm infants. *Radiology* 2014;271(1):230–238. [PubMed: 24475798]
7. Wiguna T, Guerrero AP, Wibisono S, Sastroasmoro S. Effect of 12-week administration of 20-mg long-acting methylphenidate on Glu/Cr, NAA/Cr, Cho/Cr, and mI/Cr ratios in the prefrontal cortices of school-age children in Indonesia: a study using 1H magnetic resonance spectroscopy (MRS). *Clinical neuropharmacology* 2012;35(2):81–85. [PubMed: 22318191]
8. Mueller SG, Ebel A, Barakos J, Scanlon C, Cheong I, Finlay D, Garcia P, Weiner MW, Laxer KD. Widespread extrahippocampal NAA/(Cr+Cho) abnormalities in TLE with and without mesial temporal sclerosis. *J Neurol* 2011;258(4):603–612. [PubMed: 20976465]
9. Mostert JP, Blaauw Y, Koch MW, Kuiper AJ, Hoogduin JM, De Keyser J. Reproducibility over a 1-month period of 1H-MR spectroscopic imaging NAA/Cr ratios in clinically stable multiple sclerosis patients. *Eur Radiol* 2008;18(8):1736–1740. [PubMed: 18389250]

10. Ariza M, Junque C, Mataro M, Poca MA, Bargallo N, Olondo M, Sahuquillo J. Neuropsychological correlates of basal ganglia and medial temporal lobe NAA/Cho reductions in traumatic brain injury. *Archives of neurology* 2004;61(4):541–544. [PubMed: 15096403]
11. Serles W, Li LM, Caramanos Z, Arnold DL, Gotman J. Relation of interictal spike frequency to 1H-MRSI-measured NAA/Cr. *Epilepsia* 1999;40(12):1821–1827. [PubMed: 10612351]
12. Jansen JF, Backes WH, Nicolay K, Kooi ME. 1H MR spectroscopy of the brain: absolute quantification of metabolites. *Radiology* 2006;240(2):318–332. [PubMed: 16864664]
13. Desal H, Pineda Alonso N, Akoka S. Electronic reference for absolute quantification of brain metabolites by 1H-MRS on clinical whole-body imaging. *Journal of neuroradiology Journal de neuroradiologie* 2010;37(5):292–297. [PubMed: 20334920]
14. Kirov II, Tal A, Babb JS, Herbert J, Gonen O. Serial proton MR spectroscopy of gray and white matter in relapsing-remitting MS. *Neurology* 2013;80(1):39–46. [PubMed: 23175732]
15. Kantarci K, Jack CR, Jr. Neuroimaging in Alzheimer disease: an evidence-based review. *Neuroimaging Clin N Am* 2003;13(2):197–209. [PubMed: 13677801]
16. Majos C, Alonso J, Aguilera C, Serrallonga M, Perez-Martin J, Acebes JJ, Arus C, Gili J. Proton magnetic resonance spectroscopy ((1)H MRS) of human brain tumours: assessment of differences between tumour types and its applicability in brain tumour categorization. *Eur Radiol* 2003;13(3):582–591. [PubMed: 12594562]
17. Li BS, Wang H, Gonen O. Metabolite ratios to assumed stable creatine level may confound the quantification of proton brain MR spectroscopy. *Magn Reson Imaging* 2003;21(8):923–928. [PubMed: 14599543]
18. Veenith TV, Mada M, Carter E, Grossac J, Newcombe V, Outtrim J, Lupson V, Nallapareddy S, Williams GB, Sheriff S, Menon DK, Maudsley AA, Coles JP. Comparison of inter subject variability and reproducibility of whole brain proton spectroscopy. *PLoS One* 2014;9(12):e115304. [PubMed: 25517503]
19. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143(1):29–36. [PubMed: 7063747]
20. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978;8(4):283–298. [PubMed: 112681]
21. Kirov II, Tal A, Babb JS, Lui YW, Grossman RI, Gonen O. Diffuse axonal injury in mild traumatic brain injury: a 3D multivoxel proton MR spectroscopy study. *Journal of Neurology* 2013;260(1):242–252. [PubMed: 22886061]
22. Hinkley DV. On the ratio of two correlated normal random variables. *Biometrika* 1969;56(3):635–639.
23. van Erkel AR, Pattynama PM. Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. *European journal of radiology* 1998;27(2):88–94. [PubMed: 9639133]
24. Tal A, Kirov II, Grossman RI, Gonen O. The role of gray and white matter segmentation in quantitative proton MR spectroscopic imaging. *NMR in Biomedicine* 2012;25(12):1392–1400. [PubMed: 22714729]
25. Soher BJ, Young K, Govindaraju V, Maudsley AA. Automated spectral analysis III: application to in vivo proton MR spectroscopy and spectroscopic imaging. *Magn Reson Med* 1998;40(6):822–831. [PubMed: 9840826]
26. Ao YC, Shi YB, Zhang W, Li YJ. An Approximate Calculation of Ratio of Normal Variables and Its Application in Analog Circuit Fault Diagnosis. *J Electron Test* 2013;29(4):555–565.
27. Samarakoon S, Rajatheva N, Bennis M, Latva-aho M. Outage Probability and Capacity for Two-Tier Femtocell Networks by Approximating Ratio of Rayleigh and Log Normal Random Variables. *Ieee Vts Veh Technol* 2013.
28. Celano G, Castagliola P. Design of a Phase II Control Chart for Monitoring the Ratio of two Normal Variables. *Qual Reliab Eng Int* 2016;32(1):291–308.
29. Broda S, Paoletta MS. Evaluating the density of ratios of noncentral quadratic forms in normal variables. *Comput Stat Data An* 2009;53(4):1264–1270.
30. Hinkley DV. On Ratio of 2 Correlated Normal Random Variables. *Biometrika* 1969;56(3):635–639.
31. Marsaglia G Ratios of normal variables. *J Stat Softw* 2006;16(4):1–10.

32. Pham-Gia T, Turkkan N, Marchand E. Density of the ratio of two normal random variables and applications. *Commun Stat-Theor M* 2006;35(9):1569–1591.
33. Nadarajah S Linear combination, product and ratio of normal and logistic random variables. *Kybernetika* 2005;41(6):787–798.
34. Forchini G The exact cumulative distribution function of a ratio of quadratic forms in normal variables, with application to the AR(1) model. *Economet Theor* 2002;18(4):823–852.
35. Guiard V Some Remarks on the Estimation of the Ratio of the Expectation Values of a Two-Dimensional Normal Random Variable (Correction of the Theorem of Milliken). *Biometrical J* 1989;31(6):681–697.
36. Jones MC. On Moments of Ratios of Quadratic-Forms in Normal Variables. *Stat Probabil Lett* 1987;6(2):129–136.
37. Dunlap WP, Silver NC. Confidence-Intervals and Standard Errors for Ratios of Normal Variables. *Behav Res Meth Instr* 1986;18(5):469–471.
38. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep* 1966;19(1):3–11. [PubMed: 5942109]
39. Geary RC. Testing for normality. *Biometrika* 1947;34(Pt 3-4):209–242. [PubMed: 18918691]
40. Henderson AR. Testing experimental data for univariate normality. *Clin Chim Acta* 2006;366(1-2):112–129. [PubMed: 16388793]
41. Longo R, Pollesello P, Ricci C, Masutti F, Kvam BJ, Bercich L, Croce LS, Grigolato P, Paoletti S, de Bernard B, et al. Proton MR spectroscopy in quantitative in vivo determination of fat content in human liver steatosis. *J Magn Reson Imaging* 1995;5(3):281–285. [PubMed: 7633104]
42. Wimmer M, Wilmering B, Sasse D. The relation of rat liver wet weight to dry weight. *Histochemistry* 1985;83(6):571–572. [PubMed: 4086343]

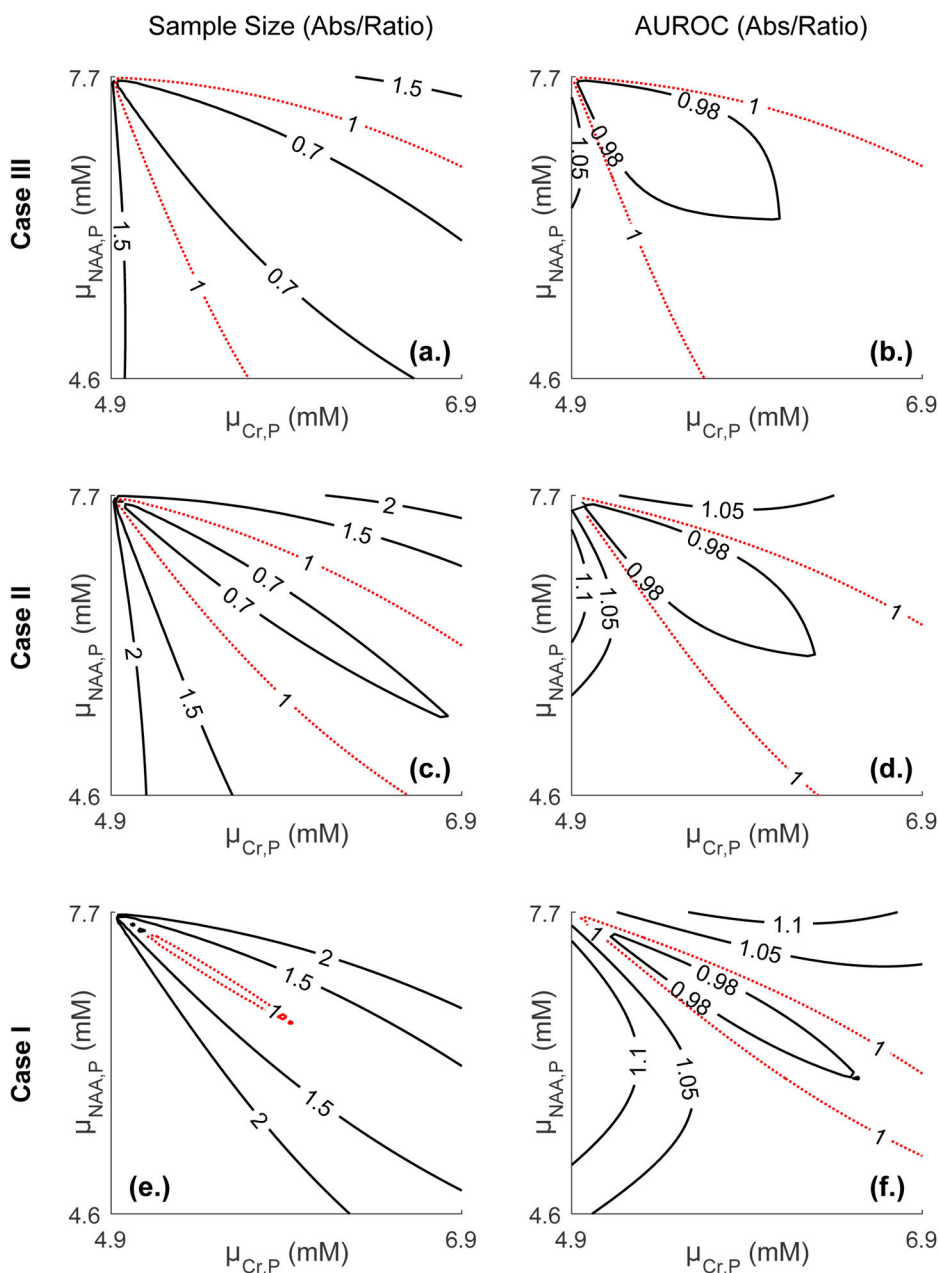


**Fig. 1.** Plots of (a.)  $h(R)$ , (b.)  $\exp(h(R))$  and (c.) the normal approximation (red) to the exact probability distribution function  $p(R)$  (blue) defined in Eq. [1], for  $\mu_1=10$ ,  $\sigma_1=2$ ,  $\mu_2=14$ ,  $\sigma_2=2$  ( $CV_2 \approx 14\%$ ). Also shown in (d.) is the function  $g(R)$ .



**Fig 2.** Receiver operating characteristic (ROC) curves: a visual overview. Shown are three pairs of normal distributions (Case I:  $4 \pm 0.5$  vs  $6 \pm 0.5$ ; Case II:  $4.7 \pm 0.5$  vs  $6 \pm 0.5$ ; Case III:  $5.4 \pm 0.5$  vs  $6 \pm 0.5$ ), possibly corresponding to patient and control distributions of some biomarker  $x$ . ROC curves for all three cases are plotted above as follows: a classifier value  $\mu_c$  is selected, with all values above it classified as “pathological” and below it as “healthy”; random values are generated from each distribution and classified, and the rate of true positives (TPR) and false positive (FPR) is recorded;  $\mu_c$  is swept from  $\mu_c=2$  to  $\mu_c=8$ , and the processes is repeated at each step, until one is able to construct a continuous curve of TPR as a function of FPR, which is the ROC curve. The dashed line corresponds to a coin toss, where cases are

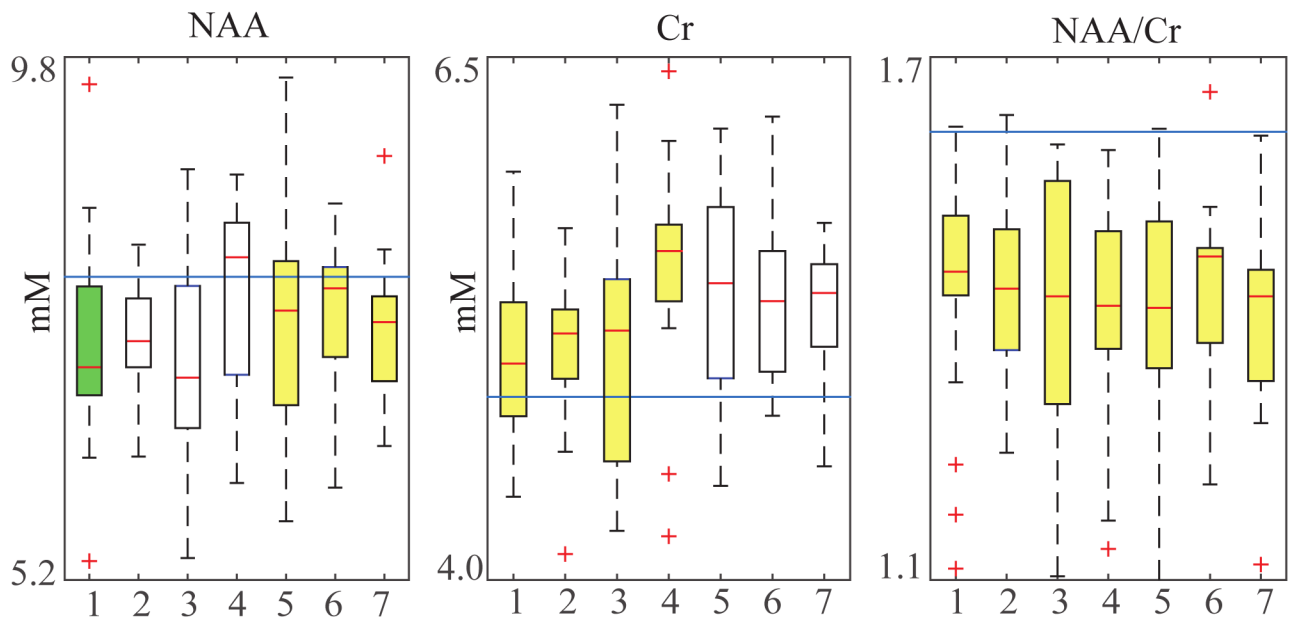
classified as pathological or healthy randomly with equal probability. As the distance between the two distributions decreases, classification becomes more prone to error and the ROC curve becomes more diagonal-like. The areas under the ROC curves are AUROC=0.99, 0.97 and 0.80 for cases I, II and III, respectively, reflecting how classification becomes more difficult as the distributions become closer (AUROC=0.5 for the random coin toss).



**Fig. 3.** Contour plots comparing sample sizes and AUROC curves between absolute quantification and ratios in the  $(\mu_{NAA, patients}, \mu_{Cr, patients})$  plane. Physiologically relevant metabolite values and standard deviations for averaged WM NAA ( $7.7 \pm 0.6$  mM) and Cr ( $4.9 \pm 0.5$  mM) were employed for controls. Patient metabolite values were varied by up to 40%, with NAA decreasing and Cr increasing. Three cases were examined: decreased (case I:  $\sigma_{NAA,P}=0.4$ ,  $\sigma_{Cr,P}=0.3$ ), unchanged (case II) and increased (case III:  $\sigma_{NAA,P}=0.9$ ,  $\sigma_{Cr,P}=0.9$ ) patient standard deviations. **Left:** the ratio of sample sizes  $N_{ratio}/\min(N_{NAA}, N_{Cr})$  is plotted, where  $N_{NAA}$ ,  $N_{Cr}$ ,  $N_{ratio}$ , are the sample sizes required for NAA, Cr, and their ratio, assuming statistical significance of 5% and power of 80% and a single-sided change. The unity

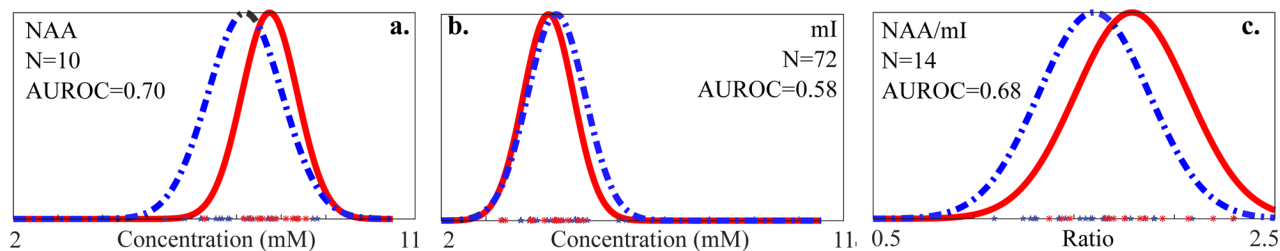


contour (red, dashed line) implies both require the same sample size. The region inside the unity contour indicates absolute quantification is preferable, while ratios are preferable outside the unity contour. **Right:** the ratio  $AUROC_{\text{ratio}}/\max(AUROC_{\text{NAA}}, AUROC_{\text{Cr}})$ . Here, again, a ratio of unity (red, dashed line) indicates both absolute quantification and ratios are equivalent in terms of their AUROC. Ratios are superior outside the unity contour.



**Fig. 4.**

Statistical significance of NAA, Cr, and NAA/Cr in longitudinally studied MS patients and controls. Data taken from (14). The straight line represents the controls' means, assumed unaltered throughout, while the boxplots represent the patient metabolite distributions. Highlighted boxes represent a statistically significant difference between controls and patients at that particular time point, as determined by a two-tailed t-test using  $\alpha=0.05$  (green and yellow) and  $\alpha=0.01$  (yellow). Note that, while neither NAA nor Cr remain statistically significant throughout, their ratio does. Population and control metabolite means and standard deviations for all time points can be found in Table 1.



**Fig. 5.**

Metabolite data from mTBI patients (blue stars) and controls (red asterisks) for (a.) global WM NAA, (b.) mI and (c.) NAA/mI, with fitted normal distributions (solid red: controls; dashed blue: patients):  $\mu_{\text{NAA}}^{(\text{controls})} = 7.74 \pm 0.62\text{mM}$ ,  $\mu_{\text{NAA}}^{(\text{patients})} = 7.19 \pm 0.83\text{mM}$ ,  $\mu_{\text{mI}}^{(\text{controls})} = 4.4 \pm 0.56\text{mM}$ ,  $\mu_{\text{mI}}^{(\text{patients})} = 4.574 \pm 0.65\text{mM}$ ,  $\mu_{\text{ratio}}^{(\text{controls})} = 1.76 \pm 0.27$ ,  $\mu_{\text{ratio}}^{(\text{patients})} = 1.57 \pm 0.29$ . NAA decreases and mI slightly increases, with the only statistically significant change being that of NAA as determined by a double-sided t-test ( $p=0.042$ ). Note the increased variability of the ratio distribution in (c), introduced by dividing by mI. The sample sizes required to observe each change (Eq. [4]) are  $N_{\text{NAA}}=10$ ,  $N_{\text{mI}}=72$ ,  $N_{\text{NAA/mI}}=14$ , and the AUROCs are  $\text{AUROC}_{\text{NAA}}=0.70$ ,  $\text{AUROC}_{\text{mI}}=0.58$ ,  $\text{AUROC}_{\text{NAA/mI}}=0.68$ , showing absolute quantification is superior to metabolic ratios in detecting mTBI WM injury.

**Table 1.**

Metabolite global WM means for NAA, Cr and NAA/Cr, over seven time points, at 6 month intervals, for both patients and control in MS (taken from (14)). This data reflects the boxplots in Fig. 4. Greyed out columns indicate statistical significance (p-value  $< 0.05$  with  $\alpha=0.05$ ). All absolute quantities are in millimoles/L of tissue water. The control values over all time points were pooled to create a global  $\mu_{\text{control}}$ ,  $\sigma_{\text{control}}$ , following (14). For each metabolite, the sample size requirement (rounded up) and AUROC are displayed. Note that, even though Cr is not statistically significant at time points 5, 6 and 7, it actually yields smaller sample size requirements and AUROC compared to NAA/Cr which is statistically significant.

Time Pt.	1	2	3	4	5	6	7
NAA ( $\mu_{\text{control}}=7.87$ , $\sigma_{\text{control}}=0.71$ )							
$\mu_{\text{patient}}$	7.26	7.30	7.14	7.70	7.50	7.58	7.39
$\sigma_{\text{patient}}$	0.94	0.53	1.02	0.82	0.90	0.61	0.61
$\lceil N \rceil$	11	9	8	131	28	35	13
AUROC $\times 100$	69	74	72	56	63	62	69
Cr ( $\mu_{\text{control}}=4.88$ , $\sigma_{\text{control}}=0.38$ )							
$\mu_{\text{patient}}$	5.09	5.09	5.09	5.49	5.37	5.34	5.30
$\sigma_{\text{patient}}$	0.46	0.39	0.63	0.54	0.51	0.41	0.31
$\lceil N \rceil$	24	19	28	4	5	5	5
AUROC $\times 100$	63	66	62	82	78	80	81
NAA/Cr ( $\mu_{\text{control}}=1.61$ , $\sigma_{\text{control}}=0.19$ )							
$\mu_{\text{patient}}$	1.43	1.43	1.40	1.40	1.40	1.42	1.39
$\sigma_{\text{patient}}$	0.23	0.15	0.26	0.20	0.21	0.16	0.14
$\lceil N \rceil$	8	7	7	6	6	6	4
AUROC $\times 100$	73	77	74	77	77	78	82