Reproducibility of the Whole-Brain $N$-Acetylaspartate Level across Institutions, MR Scanners, and Field Strengths

B. Benedetti, D.J. Rigotti, S. Liu, M. Filippi, R.I. Grossman and O. Gonen

*AJNR Am J Neuroradiol* 2007, 28 (1) 72-75

http://www.ajnr.org/content/28/1/72
Reproducibility of the Whole-Brain N-Acetylaspartate Level across Institutions, MR Scanners, and Field Strengths

BACKGROUND AND PURPOSE: Radiologic markers in multicenter trials are often confounded by different instrumentation used. Our goal was to estimate the variance of the global concentration of the neuronal cell marker N-acetylaspartate (NAA) among research centers using MR imaging scanners of different models, from different manufacturers, and of different magnetic field strength.

MATERIALS AND METHODS: Absolute millimolar amounts of whole-brain NAA (WBNAA) were quantified with nonlocalizing proton MR spectroscopy in the brains of 101 healthy subjects (53 women, 48 men) aged 16–59 years (mean, 34.2 years). Twenty-three were scanned at 1 institute in a 1.5T Siemens Vision; 31 from another institute were studied with a 1.5T Siemens SP63; 36 were scanned at a third institute (24 with a 1.5T Vision, 12 with a 3T Siemens Trio); and 11 were obtained at a fourth institute using a 4T GE Signa 5.x. The NAA amounts were quantified with phantom-replacement and divided by the brain volume, segmented from MR imaging, to yield the concentration, a metric independent of brain size suitable for cross-sectional comparison.

RESULTS: The average WBNAA concentration among institutions was 12.2 ± 1.2 mmol/L. The subjects’ WBNAA distributions did not differ significantly (p > .237) among the 4 centers, regardless of scanner manufacturer, model, or field strength and irrespective of whether adjustments were made for age or sex.

CONCLUSION: Absolute quantification against a standard makes the WBNAA concentration insensitive to the MR hardware used to acquire it. This important attribute renders it a robust surrogate marker for multicenter neurologic trials.

Human Subjects and MR Scanners

Because healthy persons are presumed, by definition, to be as similar as biologically possible, separate cohorts can be conveniently studied at different institutions, as opposed to transporting (the same) patient group around. Therefore, 101 healthy subjects (53 women, 48 men) aged 16–59 years (mean, 34.2 years) were recruited as follows: 23 subjects aged 31.0 ± 1.5 years (mean ± SD) were studied at the first institution using a 1.5T Siemens Vision (Siemens, Erlangen Ger-
many). Thirty-one subjects aged 33.0 ± 9.5 years were studied at the second institution using a 1.5T Siemens SP63. Thirty-six subjects aged 34.0 ± 8.9 years were studied at the third institution, as follows: 24 in a 1.5T Vision scanner and 12 in a 3T Siemens Trio. Finally, 11 subjects aged 43 ± 15 years were imaged at the fourth institution in a 4T Signa 5.0 (GE Medical Systems, Milwaukee, Wis). All gave written informed consent, and the study was approved by the respective Institutional Review Boards.

**MR Imaging—Brain volume, \( V_p \) Segmentation**

Each subject’s brain parenchymal volume, \( V_p \), was obtained from high-resolution T1-weighted sagittal magnetization-prepared rapid gradient-echo (TE/TR/TI, 7.0/14.7/300 ms; 128 1.5-mm sections; matrix, 256 × 256; FOV, 210 × 210-mm²). The images were segmented into tissue and CSF, as described previously, with only the parenchyma accounted for in \( V_p \).

**MR Spectroscopy—WBNAA quantification**

The amount of WBNAA, \( Q_{\text{NAA}} \), was measured in each institution by using their imager in its standard neuro-MR configuration with a transmit-receive head coil. In each case, shimming yielded consistent 12 ± 3, 20 ± 5, and 26 ± 7 Hz whole-head water linewidths at 1.5-, 3-, and 4T, respectively, followed by the same nonlocalizing TE/TR/TE = 0/970/10,000 ms WBNAA ¹H-MR spectroscopy sequence. The sequence nulls the NAA signal intensity every second shot with inversion-recovery, whereas the lipids’ signals, with their short T1, -220 ms, are always thermal (at either the TR of 10 seconds, or the TI of 970 ms). Therefore, an add-subtract scheme destructively interferes with their signal intensity, as shown in Fig 1. Absolute quantification was done against a reference 3-L sphere of 1.5 × 10⁻² mol of NAA in water. Subject and reference NAA peaks, \( S_s \) and \( S_R \), were integrated from the resultant spectrum, as shown in Fig 1, and \( Q_{\text{NAA}} \) was obtained as:

\[
1) \quad Q_{\text{NAA}} = 1.5 \times 10^{-2} \frac{S_s}{S_R} \frac{V_R^{180^\circ}}{V_R^{90^\circ}} \text{ moles,}
\]

where \( V_R^{180^\circ} \) and \( V_R^{90^\circ} \) are the transmitter voltages into 50 Ω for non-selective 1-ms inversion pulses on the reference and subject, respectively, reflecting relative coil loading. Note that slight changes were made to the sequence, increase of the chemical shift selection water suppression bandwidth, and decrease in the 1331 interpulse delays, proportional to the magnetic field strength.

To account for natural variations in human head sizes, each subject’s \( Q_{\text{NAA}} \) was divided by their \( V_R \) to yield the concentration

\[
2) \quad \text{WBNAA} = \frac{Q_{\text{NAA}}}{V_R} \text{ mM,}
\]

which is independent of brain size and therefore suitable for cross-sectional comparison. The intrasubject and intersubject variability of this metric was shown previously to be better than ±7%,.

**Statistics**

Analysis of variance (ANOVA) was used to compare sites and field strengths with respect to WBNAA. In particular, the Tukey honestly significant difference (HSD) procedure (to correct for multiple comparisons) was used in the ANOVA framework to make all pairwise comparisons among sites and field strengths with respect to WBNAA while maintaining the experiment-wise type I error rate for the set of comparisons at or below the 5% level. The ANOVA used WBNAA as the dependent variable and the model included subject age as a numeric factor and subject sex and either site or field strength as classification factors.

**Results**

The distribution of WBNAA values from each of the 5 scanners at the 4 centers are shown in Fig 2. There were no significant differences between them in terms of either the raw, unadjusted SDs or the SDs after adjustment for variation in brain size across the group of subjects. The ANOVA framework was used to make all pairwise comparisons among sites and field strengths with respect to WBNAA while maintaining the experiment-wise type I error rate for the set of comparisons at or below the 5% level. The Tukey honestly significant difference (HSD) procedure was used to correct for multiple comparisons among sites and field strengths with respect to WBNAA while maintaining the experiment-wise type I error rate for the set of comparisons at or below the 5% level.

**Fig 1.** Sample whole-head ¹H-MR spectroscopy spectra from each of the MR imaging instruments used, with institution and field strength indicated. Although all the proton metabolites can be seen in the spectrum, note that because it was acquired in a whole-head, nonlocalized fashion, only the NAA peak is implicitly localized to the brain. The area of the NAA peak of the subject, \( S_s \), was integrated in each subject as shown and converted into absolute millimoles, \( Q_{\text{NAA}} \), by comparing with the NAA peak area, \( S_R \), in a reference phantom of known concentration as described by Eq. 1. Note the excellent signal-to-noise ratio of this short acquisition and the nearly total elimination of the skull’s lipid signals from obscuring the spectrum.

**Fig 2.** Box plot showing the 25%, 50% (median), and 75% quartiles (box) and ±95% (whiskers), of the distributions of the subjects’ WBNAA concentrations in each of the 4 institutions, 5 scanners, 3 field strengths, and 2 manufacturers used in this study. Note that the differences among the distributions (means and SDs) are statistically insignificant, independent of the (healthy) subjects’ age or sex, indicating that the methodology is robust to instrumentation differences.
attributable to age or sex (p > .237), as compiled in the Table. Furthermore, there were no significant differences between the mean WBNAA concentrations among the different instruments or research centers. Therefore, the invariant WBNAA concentration of healthy subjects is 12.2 ± 1.2 mmol/L.

Tukey HSD indicated that there were no significant differences (p > .27) between any pair of sites with respect to mean level of WBNAA. Furthermore, the same test has shown no significant difference between any 2 field strengths (HSD adjusted p > .12).

Discussion

Because of their exquisite sensitivity to brain abnormalities, MR-derived metrics are increasingly used as surrogates for efficacy in clinical trials of treatments for neurologic disorders.1,2,5,7,9,30 The problem, however, is that current clinical MR metrics do not fulfill some of the prerequisites needed to define a valid paraclinical surrogate.3,4 Key among these are that it be 1) simple to implement, 2) reproducible, 3) clinically meaningful, 4) pathologically specific, 5) sensitive to disease changes over time, and 6) of reasonably low labor intensity. The goals of this study were to determine the influence of center and hardware on WBNAA, an issue critical to the consideration of this technique for inclusion in multicenter clinical trials.

Prerequisite 1, above, is met by the WBNAA technique, as demonstrated by implementation in this study, across the multiple centers and scanners of different vintage, manufacturer, and magnetic field strength. These implementations also reflect, in a more subtle way, that the conjugate requirement 6), that of low labor intensity, is also met; ie, WBNAA acquisition and postprocessing are no more time-consuming than other quantitative MR imaging metrics. These 2 are “baseline” conditions for any technique to be considered for use in any multicenter trial.

Reproducibility, requirement 2) above, is also satisfied, as shown by Fig 2, as well as the statistical analyses in the Results, summarized in Table 1. Specifically, neither the 12.2 mmol/L average mean of the WBNAA distributions, nor its ± 1.2 mmol/L average SD, differs among imagers or centers. In other words, the variability of WBNAA across multiple imagers, ±10%, is similar to that of any single instrument, assessed previously at ≥6%–7% in healthy control subjects.18,23 This finding is of particular importance if WBNAA is to be considered an adjunctive treatment efficacy measure for 2 main reasons: first, the scanners used represent many clinical research centers worldwide; second, most trials have treatment and control arms. Therefore, knowledge of the intraintstrument and interinstrument variability in the control subjects will assist in determining the subject population required to detect a given expected treatment effect in the patients.

Previous studies have shown that the scan-rescan variability is a serious issue for quantities derived from conventional13,14 and MT MR imaging.15 This has led to the formulation of rules standardizing acquisition across multiple centers, which suggested that upgrades or changes to scanners be limited during trials.16,17 This is a difficult rule to follow considering 1) the 2–3-year course of most clinical neurology trials versus the 5-year life span of modern MR scanners and 2) the manufacturers’ annual software upgrade cycles, some mandatory for regulatory compliance and others to keep the system current for its prime mission of clinical diagnostic MR imaging. The use of a calibration reference for the WBNAA measurement shows its potential to preempt both concerns.

The WBNAA method, however, is not without its limitations that need to be considered as part of a study design. Its main weakness is the total loss of localization. As a result, it is insensitive to changes smaller than 5%–6% (ie, oblivious to focal disorders [eg, stroke] and unable to discern between several foci of great loss and diffuse, overall smaller NAA decline). It also misses the NAA signal intensity from regions that suffer extreme susceptibility-driven field inhomogeneities (eg, the inferior frontal lobes, temporal lobes, and around the auditory canals). This loss, however, has been quantified previously at less than 10%.25

In conclusion, this study indicates that the use of different MR scanners should not be considered a source of confounding variability for use of WBNAA in multicenter studies of neurologic disorders. Using it, therefore, allows the study designers to focus on the much more relevant criterion of patient accrual capability rather than hardware compatibility. The premise that this approach allows an overall assessment of neuroaxonal viability makes WBNAA an attractive surrogate marker for large-scale studies of treatment outcomes, especially for conditions characterized by diffuse or widespread brain damage, (eg, multiple sclerosis, AIDS, traumatic brain injury, Alzheimer disease, normal aging).

References


---

**Table 1**: Estimated mean ± SD NAA concentrations at each institution.

<table>
<thead>
<tr>
<th>Institution</th>
<th>Subjects</th>
<th>Scanner Field, Manufacturer, Model</th>
<th>Mean WBNAA (mmol/L)</th>
<th>SD (Unadjusted) (mmol/L)</th>
<th>SD (adjusted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>n = 23</td>
<td>1.5 T, Siemens, Vision</td>
<td>11.7</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>Second</td>
<td>n = 31</td>
<td>1.5 T, Siemens, SP63</td>
<td>12.3</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Third</td>
<td>n = 24</td>
<td>1.5 T, Siemens, Vision</td>
<td>12.6</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>n = 12</td>
<td>3.0 T, Siemens, Trio</td>
<td>11.7</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Fourth</td>
<td>n = 11</td>
<td>4.0 T, GE, Signa 5.x</td>
<td>12.8</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Total</td>
<td>n = 101</td>
<td></td>
<td>12.2*</td>
<td>1.2*</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Note.—WBNAA indicates whole-brain N-acetylaspartate. The SDs are presented as both unadjusted and adjusted with respect to confounding effects from age and gender (differences were in second decimal digit or smaller).

* Average of all subjects at all institutions.