Multicenter phase II clinical trials for neurologic disorders frequently rely on surrogate markers to provide outcome measures sooner and with many fewer patients than clinical assessments.\(^1\)\(^-\)\(^4\) Such surrogates are often clinical MR imaging metrics, because of their sensitivity to disease-related abnormalities and their changes over time.\(^5\)\(^-\)\(^7\) Still, despite its sensitivity, MR imaging is not specific to the nature or true extent of brain tissue damage.\(^8\)\(^-\)\(^10\) Therefore, quantitative MR techniques (such as volumetry, magnetization transfer [MT], diffusion-weighted functional imaging, and proton MR spectroscopy \([1H-MR spectroscopy\]) are sometimes also used.\(^7\)\(^-\)\(^8\)

Although they could all potentially complement MR imaging,\(^5\)\(^-\)\(^7\) their interscanner variability, which is critical to avoid confounds in the data-consolidating stage, has not yet been fully established.\(^10\)

One of the most prominent candidates for a neurologic surrogate marker is the amino acid derivative \(N\)-acetylaspartate (NAA), the second most abundant amino acid in the mammalian brain.\(^11\)\(^-\)\(^14\) Almost exclusive to neurons and their processes,\(^15\)\(^-\)\(^17\) its level is considered to reflect their health and attenuation.\(^13\) Because NAA can be quantified noninvasively with \(1H-MR spectroscopy\), this “marker + technique” duo is ideally suited to probe the underlying biochemistry of neurologic disorders,\(^13\) both as structural abnormalities and as MR-occult pathologies.\(^12\)

Unfortunately, brain \(1H-MR spectroscopy\) so far employs either small (1–8 cm\(^3\), single voxel) or large (\(\leq 100 \text{ cm}^3\)) 2D volumes-of-interest (VOI) placed away from the skull to avoid lipid contamination, missing most of the cortex.\(^18\) These limited sizes account for less than 10% of the brain volume, at best.\(^19\)\(^-\)\(^21\) Because many neurologic disorders are diffuse, missing 90%–99% of the brain implies that the overall extent of neuronal damage is represented by NAA decline in the VOI.\(^22\) Interpretation of \(1H-MR spectroscopy\) is further complicated by longitudinal VOI repositioning errors and the use of MR imaging scanners of various makes, model, and field strengths. Although the repositioning and partial coverage limitations have recently been overcome with nonlocalizing \(1H-MR spectroscopy\) quantification of the whole-brain NAA (WB-NAA) concentration,\(^23\)\(^-\)\(^25\) its sensitivity to instrumentation change has not yet been established. This issue is examined in the present article by comparing the distribution of WB-NAA from several research centers and using various imagers by 2 manufacturers operating at different magnetic field strengths.

### Materials and Methods

#### 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 (33 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-
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.x (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-mm\(^2\)). The images were segmented into tissue and CSF, as described previously,\(^{26-28}\) 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/TI = 0/970/10,000 ms WBNAA \( ^1 \)H-MR spectroscopy sequence.\(^{23}\) The sequence nulls the NAA signal intensity every second shot with inversion-recovery, whereas the lipids’ signals, with their short T1, \(~\sim 220\) 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,\(^{23}\) as shown in Fig 1. Absolute quantification was done against a reference 3-L sphere of \( 1.5 \times 10^{-2} \) 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\(^{23,28}\):

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

where \( V_R^{180 \degree} \) and \( V_s^{180 \degree} \) 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,\(^{23}\) 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 \( V_p \) to yield the concentration

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

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%,\(^{18,23}\)

**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, un-adjusted SDs or the SDs after adjustment for variation.
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 intrainstrument 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 conventional11,32 and MT MR imaging.33 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.5,31 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.** Mean WBNAA (mmol/L) and its standard deviation (SD) (unadjusted and adjusted for age and gender) at each institution.

<table>
<thead>
<tr>
<th>Institution</th>
<th>Subjects</th>
<th>Scanner Field, Model, Manufacturer</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>Fourth</td>
<td>n = 11</td>
<td>3.0 T, Siemens, Trio</td>
<td>11.7</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Total</td>
<td>n = 101</td>
<td>4.0 T, GE, Signa 5.6</td>
<td>12.8</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

* Average of all subjects at all institutions.

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).