Discrimination between Neoplastic and Nonneoplastic Brain Lesions by Use of Proton MR Spectroscopy: The Limits of Accuracy with a Logistic Regression Model

Jennifer Butzen, Robert Prost, Veerappu Chetty, Kathleen Donahue, Ronald Neppl, William Bowen, Shi-Jiang Li, Victor Haughton, Leighton Mark, Thomas Kim, Wade Mueller, Glenn Meyer, Hendrikus Krouwer and Scott Rand

AJNR Am J Neuroradiol 2000, 21 (7) 1213-1219
http://www.ajnr.org/content/21/7/1213

AJNR Am J Neuroradiol 21:1213–1219, August 2000

Discrimination between Neoplastic and Nonneoplastic Brain Lesions by Use of Proton MR Spectroscopy: The Limits of Accuracy with a Logistic Regression Model

Jennifer Butzen, Robert Prost, Veerappu Chetty, Kathleen Donahue, Ronald Neppl, William Bowen, Shi-Jiang Li, Victor Haughton, Leighton Mark, Thomas Kim, Wade Mueller, Glenn Meyer, Hendrikus Krouwer, and Scott Rand

BACKGROUND AND PURPOSE: **The most accurate method of clinical MR spectroscopy (MRS) interpretation remains an open question. We sought to construct a logistic regression (LR) pattern recognition model for the discrimination of neoplastic from nonneoplastic brain lesions with MR imaging–guided single-voxel proton MRS data. We compared the LR sensitivity, specificity, and receiver operator characteristic (ROC) curve area (Az) with the sensitivity and specificity of blinded and unblinded qualitative MRS interpretations and a choline (Cho)/*N*-acetylaspartate (NAA) amplitude ratio criterion.**

METHODS: **Consecutive patients with suspected brain neoplasms or recurrent neoplasia referred for MRS were enrolled once final diagnoses were established by histopathologic examination or serial neurologic examinations, laboratory data, and imaging studies. Control spectra from healthy adult volunteers were included. An LR model was constructed with 10 input variables, including seven metabolite resonance amplitudes, unsuppressed brain water content, water line width, and the final diagnosis (neoplasm versus nonneoplasm). The LR model output was the probability of tumor, for which a cutoff value was chosen to obtain comparable sensitivity and specificity. The LR sensitivity and specificity were compared with those of qualitative blinded interpretations from two readers (designated A and B), qualitative unblinded interpretations (in aggregate) from a group of five staff neuroradiologists and a spectroscopist, and a quantitative Cho/NAA amplitude ratio $>$ 1 threshold for tumor. Sensitivities and specificities for each method were compared with McNemar's chi square analysis for binary tests and matched data with a significance level of 5%. ROC analyses were performed where possible, and Az values were compared with Metz's method (CORROC2) with a 5% significance level.**

RESULTS: **Of the 99 cases enrolled, 86 had neoplasms and 13 had nonneoplastic diagnoses. The discrimination of neoplastic from control spectra was trivial with the LR, reflecting high homogeneity among the control spectra. An LR cutoff probability for tumor of 0.8 yielded a specificity of 87%, a comparable sensitivity of 85%, and an area under the ROC curve of 0.96. Sensitivities, specificities, and ROC areas (where available) for the other methods were, on average, 82%, 74%, and 0.82, respectively, for readers A and B, 89% (sensitivity) and 92% (specificity) for the group of unblinded readers, and 79% (sensitivity), 77% (specificity), and 0.84 (Az) for the Cho/NAA $>$ 1 criterion. McNemar's analysis yielded significant differences in sensitivity (n$\sim$86 neoplasms) between the LR and reader A, and between the LR and the Cho/NAA $>$ 1 criterion. The differences in specificity between the LR and all other methods were not significant (n$\sim$13 nonneoplasms). Metz's analysis revealed a significant difference in Az between the LR and the Cho/NAA ratio criterion.**

Received August 27, 1999; accepted after revision January 20, 2000.

From the Departments of Radiology, Biophysics (J.B., R.P., V.C., K.D., R.N., W.B., S-J.L., V.H., L.M., T.K., W.M., G.M., H.K., S.R.), Neurosurgery (W.M., G.M.), and Neurology (H.K.), Medical College of Wisconsin, Milwaukee, Wisconsin; the Department of Family Practice (V.K.C.), University of Illinois College of Medicine at Rockford, Rockford, Illinois; Department of Radiology (V.H.), University of Wisconsin Hospitals and Clinics, Madison, Wisconsin; and the Department of Radiology (T.K.), University of Washington Medical Center, Seattle, Washington.

Address reprint requests to Scott D. Rand, M.D., Ph.D., Department of Radiology, Medical College of Wisconsin, Froedtert Memorial Lutheran Hospital, 9200 West Wisconsin Avenue, Milwaukee, Wisconsin 53226.

CONCLUSION: **The upper limits of sensitivity, specificity, and ROC area achieved in the construction of the LR model with MRS data demonstrate the potential for improved discrimination of neoplasm from nonneoplasm relative to either qualitative MRS interpretation by blinded readers or by quantitative interpretation with a Cho/NAA amplitude ratio threshold. The sensitivity, specificity, and ROC curve area of the LR were comparable to unblinded MRS readers who had the benefit of prior imaging studies and clinical data.**

Proton MR spectra obtained from brain neoplasms typically show: 1) decreased *N*-acetylaspartate (NAA), a marker of neuronal integrity, 2) diminished Creatine (Cr), involved in cellular energetics and osmotic balance, and 3) increased Choline (Cho), involved in cell membrane turnover. Lactate (Lac) and mobile lipids (Lip) can be evident in aggressive tumors, reflecting increased anaerobic metabolism and cellular necrosis, respectively. Recently, qualitative (1) and quantitative (2) interpretations of single-voxel spectra have been used to discriminate neoplastic from nonneoplastic brain lesions. However, the most accurate means of clinical MR spectroscopy (MRS) interpretation remains an open question. Qualitative methods (1), thresholds for metabolite ratios (2), and statistical pattern recognition techniques such as logistic regression (LR), linear discriminant analysis (LDA), (3) and neural networks have been applied to MRS interpretation. The purpose of this study was to construct a statistical LR model designed to distinguish neoplastic lesions from either normal brain parenchyma or from nonneoplastic brain lesions, and to compare the upper limits of sensitivity and specificity of the LR model with that of blinded readers, unblinded readers, and a Cho/NAA amplitude ratio threshold. An LR model constructed with inputs including the final diagnosis (neoplasm vs nonneoplasm) is capable of prospectively predicting the probability of neoplasm in unknown cases.

The LR and other multivariate techniques have advantages over qualitative interpretations because: 1) all resonances can be used simultaneously and consistently, 2) the decision rule for a positive test can be adjusted explicitly to tailor the trade-off between sensitivity (type I errors) and specificity (type II errors), 3) models can retain and use all information from a large number of prior cases, and 4) separate LR models can be constructed for different classes of patients such as those with treated versus untreated lesions. Potential disadvantages of the LR include the additional computational resources required beyond spectral postprocessing to implement and automate the method and the need to maintain a quantitative clinical MRS database that includes final diagnoses.

Methods

Study Subjects

Consecutive patients with suspected brain neoplasms on CT or MR scans studied with MRS over an approximate 2-year period were included after a final diagnosis (either neoplastic or nonneoplastic) was established either by histologic examination of a biopsy specimen, or by serial imaging studies, laboratory tests, and clinical course. Patients with chronic temporal lobe epilepsy (TLE) and working diagnoses of mesial temporal sclerosis versus low-grade glioma constituted one subgroup. Several patients were excluded owing to corruption of the MRS archive that precluded the measurement of resonance amplitudes. MRS was also performed in the left frontal lobe of eight healthy adult control subjects. Separate LR models were constructed to distinguish neoplasms in patients from either normal cerebrum in control subjects or from abnormal, nonneoplastic lesions in patients. In distinguishing neoplasms from normal brain or from nonneoplastic brain lesions, separate LR models were constructed with raw resonance amplitudes (arbitrary units) obtained with commercially available software (GE/SAGE, GE Medical Systems, Waukesha, WI), or with amplitudes normalized by the MRS voxel size (arbitrary units per cc).

MRS Technique

MR imaging–guided single-voxel proton MRS was performed on a clinical 0.5-T system with the point-resolved spectroscopy (PRESS) pulse sequence (1500 /41–256 [TR/TE]), chemical-shift selective (CHESS) water suppression, and conventional postprocessing techniques described elsewhere (4). Localizer images and water-suppressed and unsuppressed spectra were typically acquired within 45 minutes (1). Patient spectra were acquired with a prototype quadrature receive/transmit head coil or a receive-only conformal surface coil (Medical Advances Inc., Milwaukee, WI). Cubic or nearly cubic MRS voxels of 1–3 cc were centered over solid portions of the lesions to sample the most metabolically active tissue, and to avoid necrotic debris or edema whenever possible. Regions that revealed postcontrast enhancement on previous studies were sampled whenever possible. Spectra were obtained from the right and left hippocampi of patients with TLE.

Control spectra were acquired from 8-cc voxels with a standard head coil. All other hardware, pulse sequence parameters, and postprocessing techniques were identical to those of the patient examinations. For each control spectrum, a mixture of cortex and subcortical white matter at the same location within the posterior left frontal lobe was sampled at the level of the lateral ventricles, as determined by axial T2-weighted localizer images.

LR Model Inputs

A commercially available LR model (Stata Corporation, College Station, TX) was constructed to classify spectra as neoplastic or nonneoplastic. For each case, inputs (explanatory variables) to the LR model included: 1) the final diagnosis (either neoplastic or nonneoplastic) as the binary dependent variable, 2) up to seven brain metabolite resonance amplitudes, 3) water content (H_2O content) expressed as the area under the Lorentz-fitted unsuppressed water resonance (arbitrary units per cubic millimeter of brain), and 4) the Lorentz-fitted unsuppressed water line width (H_2O lw, full width at half maximum, expressed in Hz). The LR model calculated the probability of neoplasm as the output.

In constructing LR models to distinguish control spectra from neoplasms, resonance amplitudes from NAA (2.0 ppm), the combination (Glx, maximum within 2.2–2.4 ppm) of glu-

TABLE 1: Single-voxel proton MRS for brain neoplasia vs nonneoplasia

Method of Interpretation	*n*	Sensitivity	Specificity	ROC Area (Az)
LR (raw amplitude inputs, cutoff = 0.8)	99	87%	85%	0.96
Blinded reader A	86	75%*	90%	0.81
Blinded reader B	90	88%	58%	0.82
Average of A & B		82%	74%	0.82
Unblinded readers	95	89%	92%	NA
Cho:NAA > 1 positive for tumor	99	79%*	77%	0.84†

Note.—n signifies the number of cases with spectra considered of diagnostic quality for each method of interpretation. NA = not available.

* Signifies a statistically significant difference relative to the LR by McNemar's Chi Square method ($P \leq 0.05$).

† Signifies a statistically significant difference relative to the LR by Metz's method (CORROC2) ($P \leq 0.05$).

TABLE 2: LR input variables and correlation coefficients (n = 99 cases [86 neoplastic, 13 nonneoplastic])

LR Input Variable	LR Input Coefficient *P* Value
Lac*	0.0116
Lip*	0.0105
NAA*	0.0104
Cho*	0.0077
Cr*	0.0561
Glx*	0.2588
m-Ins*	0.9934
H_2O content†	0.1502
H_2O lw‡	0.1192

* Raw resonance amplitude (arbitrary units).

† Lorentz-fitted H_2O area per voxel size (arbitrary units × ppm/cc).

‡ Lorentz-fitted H_2O linewidth, full width at half-maximum (Hz).

tamate (Glu) and glutamine (Gln), Cr and phosphocreatine (Cr, 3.0 ppm), Cho-containing compounds (Cho, 3.2 ppm), and myo-Inositol (mIns, 3.5 ppm) were used. Lip (maximum of 0.9, 1.3 ppm) and Lac (doublet 1.15 and 1.5 ppm) resonances were not detected on control spectra, and were not included as inputs.

In constructing models to distinguish nonneoplastic from neoplastic lesions in patients, Lip and Lac resonance amplitudes were included. The Lip resonance was defined as the maximum at either 0.9 ppm (methyl) or 1.3 ppm (methylene) within the broad resonance. Overlapping Lip and Lac peaks were not further separated by spectra acquired with a long TE value of 272 milliseconds. The Lac resonance at 1.15 ppm was used, and the complementary 1.5-ppm peak in the lactate doublet (centered at 1.3 ppm) was excluded when it made little difference in the LR model output.

LR Model Output

The LR model computed the probability of neoplasm ranging from 0 to 1 as the output. The cutoff value (cutoff probability) for a positive MRS examination for neoplasia was adjusted to obtain equal rates of false-negative type II (relating to sensitivity) and false-positive type I (relating to specificity) errors. A receiver operator characteristic (ROC) curve for the model output was generated and the area under the curve was calculated (Table 1). ROC results generated with Stata software (Stata Corporation, College Station, TX) were corroborated with labroc1 software (C. E. Metz, University of Chicago). Logistic regression coefficients for the explanatory variables, including brain metabolite amplitudes, water content, and water line width, were considered statistically significant when *P* values were less than or equal to .05 (Table 2).

Qualitative Interpretations and the Cho/NAA Ratio

Qualitative interpretations were made by two blinded neuroradiologists (readers A and B), and by one of five unblinded staff neuroradiologists plus a staff spectroscopist. Blinded readers classified the control and patient spectra as diagnostic or not, and, if diagnostic, as neoplastic (with a score from 50 to 100) or nonneoplastic (with a score from 1 to 49) when abnormal. With the benefit of the spectroscopist's opinion and prior imaging studies, patient history, and laboratory results provided by the referring physician, the (unblinded) staff neuroradiologist interpreted the spectra as diagnostic or not, and, if diagnostic, as neoplastic or nonneoplastic (binary score). Blinded and unblinded readers were given the discretion to declare a spectrum as nondiagnostic if the technical quality was insufficient or if the findings were equivocal. Some initial nondiagnostic examinations were repeated at the same or a different location. This resulted in different sample sizes, n, for each reader (Table 1). MR spectra were also interpreted quantitatively such that a Cho/NAA amplitude ratio greater than unity was considered positive for neoplasm (2).

For the discrimination of neoplasms from nonneoplasms in patient spectra, maximum likelihood estimates of binormal ROC curves were performed for interpretations by the blinded readers, and by the Cho/NAA threshold (labroc1 software). The area under the ROC curve for reader B was determined graphically from discrete operating points when the estimation algorithm did not converge to provide a continuous curve after 100 iterations. Unblinded interpretations were made with a binary score, thus precluding a ROC analysis.

Statistical Comparison of Methods of MRS Interpretation

Pair-wise differences in sensitivity and specificity between the LR model and the blinded readers, the unblinded readers, and the Cho/NAA ratio criterion were examined with McNemar's Chi-square method (5) for matched, potentially correlated dichotomous (binary) tests (InStat, GraphPad Software, San Diego, CA; Stata software, Stata Corporation, College Station, TX). Differences in sensitivity and specificity were considered signficant when *P* values were less than or equal to .05. Pair-wise differences in the area under the ROC curve, Az, between the LR model and the other interpretations were examined with CORROC2 software (C. E. Metz) designed for matched, potentially correlated datasets (6). The LR output, blinded interpretations, and Cho/NAA ratios were converted to an ordinal rating scale of up to 11 categories for use with CORROC2. Differences in ROC curve area were considered signficant when *P* values were less than or equal to .05. The small, unbalanced study sample with a greater number of neoplastic than nonneoplastic cases precluded an analysis for statistically significant differences in Az with the Wilcoxon technique (7).

Results

Study Subject Demographics and Diagnoses

Ninety-nine patient spectra were recorded in patients in whom final diagnoses were available. Fifty-eight lesions were sampled from male and 41 from female patients. Patient age ranged from 14 to 81 years, with a mean age of 42 years. The final diagnosis was neoplastic for 86 lesions, and nonneoplastic for 13 lesions. Diagnoses were estab-

lished by histologic examination in 94 cases, and by serial imaging studies, clinical examination, and laboratory findings in five cases.

The neoplastic spectra included the following diagnoses: 27 glioblastomas, five astrocytomas grade III, seven astrocytomas grade II, eight astrocytomas grade I, seven mixed gliomas, three oligodendrogliomas, two ependymomas, one medulloblastoma, two dysembryoblastic neuroectodermal tumors, one giant cell astrocytoma, one esthesioneuroblastoma, one lymphoma, 11 meningiomas, and 10 mestastases. Nineteen of the 86 neoplasms had received prior surgical, radiation, or chemotherapeutic treatment. The nonneoplastic spectra included the following diagnoses: three ischemic infarcts, two demyelinating lesions, one sarcoidosis, one Rathke's pouch cyst, one radiation necrosis (without viable neoplasm), one arteriovenous malformation with hemorrhage, one gliosis, one abcess, one cortical dysplasia, and one herpes viral encephalitis.

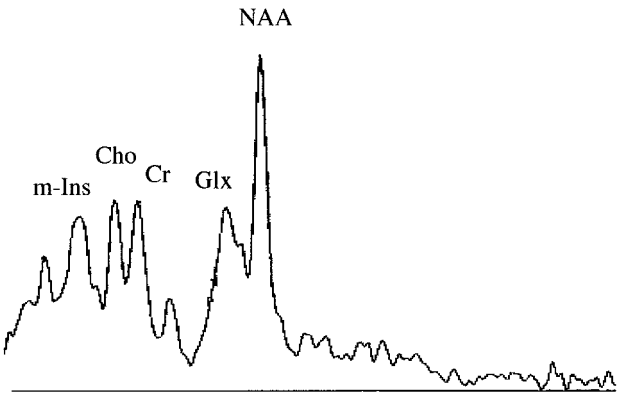


FIG 1. A single-voxel (8-cc) proton spectrum obtained from a mixture of cortex and subcortical white matter within the posterior left frontal lobe of a healthy adult volunteer was acquired with the PRESS pulse sequence (TR = 1500 / TE = 41 / 256 averages) using CHESS water suppression at 0.5 T. The NAA resonance amplitude is approximately twice that of Glx, Cr, Cho and m-Ins.

Neoplastic versus Control Spectra

With the normalized metabolite amplitudes as inputs, the probability of neoplasm (LR model output) was 1.00 for each of the 86 neoplastic spectra. For the eight control spectra, the probability of neoplasm ranged from 4.76×10^{-24} to 2.86×10^{-8}. For practical purposes, the probability was 0 for each of the control spectra. The nominal (default) cutoff probability of 0.5 for a positive MRS examination for neoplasm required no adjustment to render the sensitivity comparable to the specificity, because the LR model classified all of the spectra correctly for any cutoff value between 0 and 1. The area under the ROC curve was 1.0, indicating flawless separation of the control spectra from the neoplasms. Similarly, all of the 94 spectra were classified correctly as normal or neoplastic when the LR model used the raw metabolite amplitudes as inputs.

Neoplastic versus Nonneoplastic Patient Spectra

The sensitivity, specificity, and area under the ROC curve in the differentiation of neoplasm from nonneoplasm were 87%, 85%, and 0.96 for the LR (using raw metabolite amplitudes and a cutoff probability of 0.8), 82%, 74%, and 0.82 for the average of blinded readers A and B, 89% and 92% for the unblinded readers (Az not available), and 79%, 77%, and 0.84 for the Cho/NAA ratio criterion, respectively (Table 1). Differences in sensitivity between the LR and reader A (n=76 neoplastic cases) and between the LR and the Cho/NAA ratio (n=86 neoplastic cases) were significant. Differences in specificity between the LR and all other methods (n=10 to 13 nonneoplastic cases) were not significant. The difference in ROC area between the LR and the Cho/NAA ratio was significant, whereas the differences between the LR and the blinded and unblinded readers were not significant.

Examples of spectra from normal cerebrum, a neoplastic lesion, and a nonneoplastic lesion are illustrated in Figures 1, 2, and 3, respectively. A scatter plot of the LR probability of neoplasm based on raw metabolite amplitudes is presented in Figure 4. The corresponding probability distributions for neoplastic and nonneoplastic spectra are presented as histogram (frequency) plots in Figure 5. There were 75 true-positive, two false-positive, 11 true-negative, and 11 false-negative findings. The false-positive findings included one case of radiation necrosis without viable tumor in which Lip or Lac or both dominated the spectrum (1) and one untreated case of herpes viral encephalitis (type HHV6) with an extensive white blood cell infiltrate (8). The 11 false-negative results included: four glioblastomas, all of which had low Cho resonances (9) and three of which had Lip peaks comparable to or dominating other metabolites; four meningiomas, three of which had low Cho peaks and one of which had prominent Lip; two metastases that resembled the glioblastomas; and one mixed glioma.

Normalizing the brain metabolite amplitudes with respect to the voxel size had a small influence upon the LR model. With normalized metabolite amplitudes as inputs and a cutoff probability of 0.8, two true-positive results were converted to false-negative results, and five false-negative results were converted to true-positive results. In the aggregate, the number of false-negative findings decreased from 11 to eight, and the number of true-positive findings was increased from 75 to 78. The sensitivity changed four percentage points from 87% to 91% and the specificity remained unchanged at 85%. The area under the ROC curve changed minimally from 0.96 to 0.95.

P values for the LR coefficients corresponding to each explanatory variable are listed in Table 2.

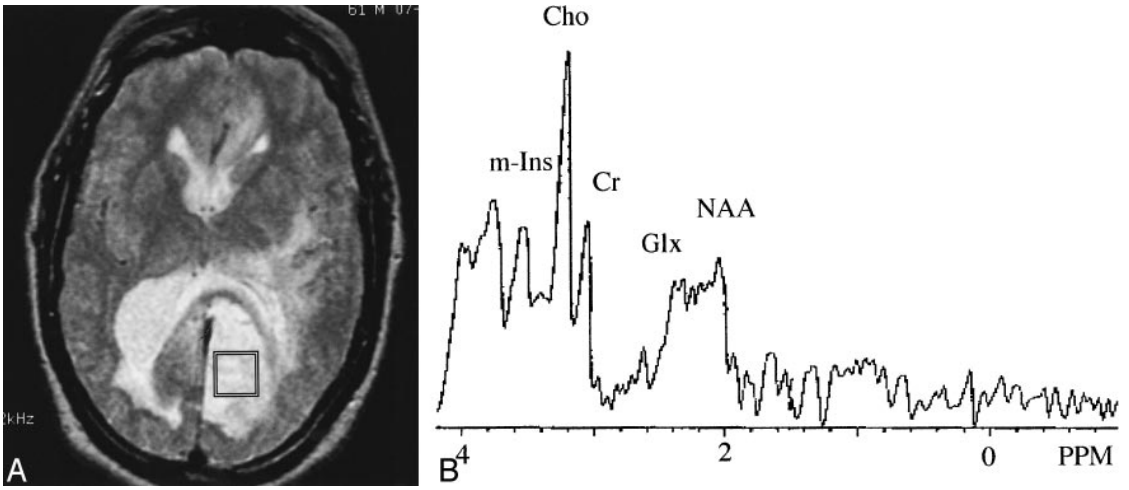


FIG 2. A CT scan of a 61-year-old man with a remote history of radiation therapy for a left glomus jugulare that revealed low attenuation in the left parasagittal occipital lobe was interpreted as compatible with a cortical neoplasm vs ischemia (not shown). An axial T2-weighted image (*A*) obtained from the same patient reveals thickened, hyperintense cortex with mass effect and adjacent white matter vasogenic edema that were interpreted as most compatible with a neoplasm. An MR spectrum (*B*) shows diminished NAA, elevated Cho, and no Lip or Lac that was interpreted as neoplastic by an unblinded reader, blinded readers A and B, a Cho/ NAA ratio criterion, and the LR model. Biopsy and histopathologic examination revealed a grade II astrocytoma.

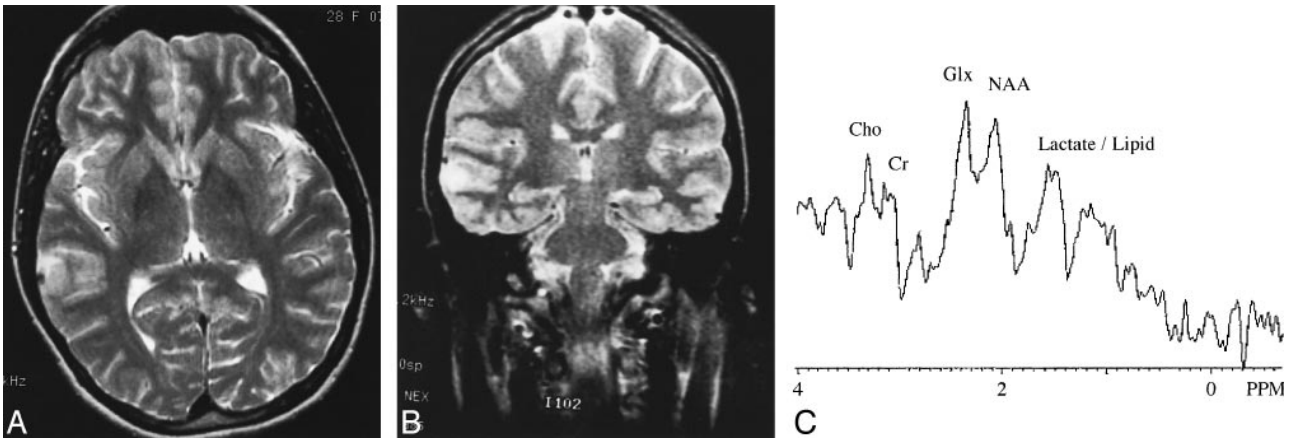
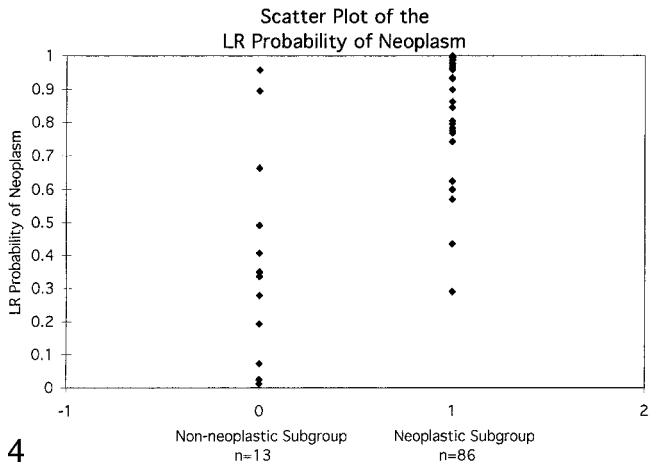


FIG 3. Axial (*A*) and coronal (*B*) T2-weighted images (3500/81 [TR/TE_{eff}]) of a 28-year-old woman with new-onset seizures that depict hyperintense, thickened cortex within the right temporal lobe were interpreted as compatible with a low-grade glioma vs an atypical cortical dysplasia. Postcontrast T1-weighted images showed no lesion enhancement (not shown). An MR spectrum (*C*) revealing elevated Glx, high Lac (doublet at 1.15 and 1.5 ppm at 0.5 T), diminshed NAA, but no elevation of Cho was interpreted as compatible with a nonneoplastic process by an unblinded reader, blinded reader B (disqualified by A), a Cho/NAA ratio criterion, and the LR model. A subsequent cerebral angiogram was normal. Biopsy and histopathologic examination revealed acute necrotizing vasculitis with recent infarction.

LR coefficients for Lac, Lip, NAA, and Cho were statistically significant, whereas coefficients for Cr, Glx, m-Ins, water content, and water line width were not significant.

Discussion

We have demonstrated that, in the construction of an LR model with the knowledge of the final diagnosis, in vivo proton MR spectra obtained from brain can be classified without error as either normal or neoplastic based on raw or normalized (by voxel volume) metabolite resonance amplitudes, water content, and water line width. The high degree of homogeneity among the control spectra contributes greatly to this success. Flawless separation of 91 untreated brain tumors from 14 control subjects was achieved with another statistical pattern recognition technique, LDA, where the LDA inputs included ratios of resonance amplitudes in neoplasms to the Cr amplitude in contralateral, unaffected brain (3). Similarly, nearly perfect results have been reported for qualitative MRS interpre-



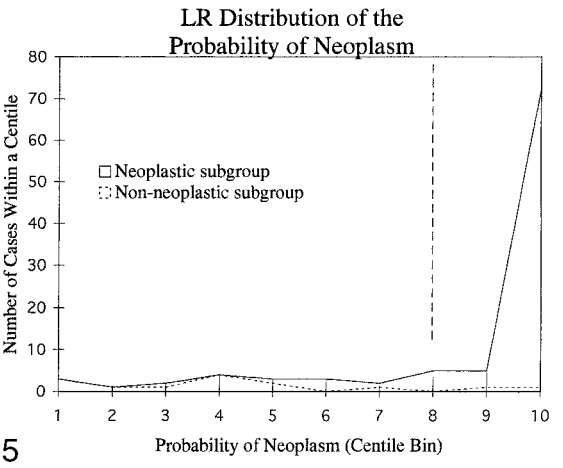


FIG 4. The LR probability of neoplasia for the neoplastic (n=86) and nonneoplastic (n=13) subgroups is presented as a scatter plot. The LR model was constructed with raw metabolite amplitudes.

FIG 5. The LR distribution of the probability of neoplasia for the neoplastic (n=86) and nonneoplastic (n=13) subgroups is plotted with an *x*-axis centile scale. The LR output, the probability of neoplasm, is constrained between 0 and 1 by definition. The 1st centile represents a bin for cases in which LR probability falls between 0 and 0.1, the 2nd centile represents a bin for LR probabilities between 0.1 and 0.2, and so on. For example, 71 neoplastic cases with LR probabilites between 0.9 and 1.0 were distributed into the 10th centile bin. The dashed vertical line over the 8th centile represents the cutoff probability of neoplasm (0.8) chosen to render the sensitivity comparable to the specificity.

tations (1), where blinded readers distinguished 10 normal from 53 abnormal spectra (both neoplastic and nonneoplastic) with an average sensitivity and specificity of 97% and 93%, respectively.

The discrimination of neoplastic from abnormal, nonneoplastic brain is clearly more challenging and clinically relevant. We are not aware of any reports in which the LR or other statistical techniques have addressed this clinical question with MRS data. Heterogeneity within and overlap between the neoplastic and nonneoplastic spectral patterns inevitably contribute to false-positive and false-negative errors. We have demonstrated that in construction of the LR model with the knowledge of the final diagnosis, the LR sensitivity exceeded that of a blinded MRS reader. Similarly, both the sensitivity and the area under the ROC curve exceeded that of a Cho/NAA ratio criterion. Demonstration of potential differences in specificity between the LR and the other interpretations would require a larger subgroup of nonneoplastic brain lesions in future studies. Demonstration of potential differences between the LR and other methods when applied specifically to the differential diagnosis of recurrent tumor versus posttreatment effects would also require a larger subgroup of treated patients.

The LR may have used Cho and NAA information more effectively than the Cho/NAA ratio criterion, and it may have benefited from other metabolite inputs. Statistically significant LR coefficients ($P \leq .05$) for Lac and Lip suggest that they were useful inputs. However, the use of traditional *P* values in multivariate regression to determine the relative contributions of each individual input is limited, because the *P* value corresponding to an input reflects the model behavior as that input is varied, keeping all other variables fixed. The assumption that all variables but one be fixed is flawed in the study of brain neoplasms, because elevated Cho with a concurrent decrease in NAA and Cr has been reported in many brain neoplasms. An analysis of the relative contributions of each input under realistic conditions would require the application of another statistical technique, the Sharpe-Markowitz decompostion, in future studies.

The group of unblinded readers had the benefit of clinical history, laboratory results, and MR findings that were unavailable to the LR, yet no significant difference in sensitivity was observed between them. This suggests the possibility for improvement of the LR in future studies through the incorporation of non-MRS data. For example, the subjective, pre-MRS probability of neoplasm according to the referring physician and to the radiologist, expressed on an ordinal scale such as 1 to 100, could be incorporated into the LR as additional explanatory variables. In this manner, diagnostic performance based on clinical examination and laboratory data obtained prior to MRS, unblinded MR interpretations alone, MRS findings alone, and any synergy between the techniques could be revealed.

To the extent that this initial series is representative of a larger referral base, these results represent an upper bound for diagnostic performance when out-of-sample predictions are made in cases in which the final diagnosis is unknown. The ''leaving one out'' (LOO) method can be used for concurrent construction and testing of the LR and other statistical models with small datasets. In the LOO method, out-of-sample predictions are made with the final diagnosis witheld for one case at a time, in succession. We elected not to pursue the LOO method, because our dataset was unbalanced,

with a minority of nonneoplastic referrals that demonstrated a broad, heterogenous probability distribution in Figure 1. Because eight of 10 nonneoplastic diagnoses in our series were represented with a single case, the LOO method would have required the LR to classify these eight cases without any prior information regarding these diseases. An alternative to the LOO method that would emulate clinical practice would be: 1) to obtain an out-of-sample prediction for new cases, and 2) to incorporate new cases into the model as soon as the final diagnoses are established by histopathologic analysis or follow-up imaging and clinical data.

A possible limitation of this study is that prior treatment in a subset of patients with a history of neoplasia could introduce changes relative to spectra obtained from untreated lesions. Such posttreatment MRS changes may render the tumor group more heterogeneous and diminish the diagnostic accuracy of the LR and the methods to which the LR was compared. Another possible limitation is that metabolite resonance amplitudes were used as LR inputs rather than absolute metabolite concentrations based on resonance areas calibrated with brain water or another (internal or external) reference. Regarding the LR input variables, the normalization of either resonance amplitudes or resonance areas by the voxel size, the unsuppressed water content, scanner settings such as receiver/transmitter amplifier gains, or a combination of factors is mathematically equivalent to the application of a multiplicative weight to each resonance in the spectrum that is unique to that spectral acquisition. However, such weights would have little relevance to the LR, because the LR captures differences between resonances within each spectrum, rather than differences in the same resonance between spectra (personal communication, Stata Corporation Technical Staff). Within the limits of spectra uncorrected for the fraction of CSF or other potential fluid collections with the MRS voxel, an insignificant *P* value for the water LR coefficient (Table 2) suggests relatively little difference in water content between solid neoplasms and nonneoplasms, assuming that all other input variables are fixed. Our findings are consistent with extensive ex vivo MRS evidence gathered in the 1970s and 1980s for which the separation of benign from malignant neoplasms based on water content, water line width (T2 lifetimes), and water T1 lifetime was generally unsuccessful (10).

Conclusion

The upper limits of sensitivity, specificity, and ROC area achieved in the construction of the LR model with MRS data demonstrate the potential for improved discrimination of neoplasm from nonneoplasm relative to either qualitative MRS interpretation by blinded readers or by quantitative interpretation with a Cho/NAA amplitude ratio threshold. The sensitivity, specificity, and area beneath the ROC curve of the LR were comparable to unblinded MRS readers who had the benefit of prior imaging studies and clinical data.

Acknowledgments

The authors wish to thank Ralph Hashoian, of Medical Advances Inc., for hardware support and Tom Raidy, of GE Medical Systems, for software support. We thank Anne Papke, Signe Haughton, Dan O'Brien, and Cathy Marszalkowski for their assistance in maintaining an MR spectroscopy database and for the preparation of spectra for blinded interpretation.

References

1. Rand SD, Prost R, Haughton V, et al. **Accuracy of single voxel proton MRS at 0.5T in neoplastic vs. nonneoplastic brain lesions.** *AJNR, Am J Neuroradiol* 1997;18:1695–1704
2. Prost R, Haughton V, Li S-J. **Brain Tumors: localized H-1 MR spectroscopy at 0.5 T.** *Radiology* 1997;204:235–238
3. Preul MC, Caramanos Z, Collins DL, et al. **Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy.** *Nat Med* 1996;2:323–325
4. Prost RW, Mark,Mewissen M, Li S. **Detection of glutamate/glutamine resonances by H-1 magnetic resonance spectroscopy at 0.5 Tesla.** *Magn Reson Med* 1997;37:615–618
5. Dwyer AJ. **Matchmaking and McNemar in the comparison of diagnostic modalities.** *Radiology* 1991;178:328–30
6. Metz CE. **Some practical issues of experimental design and data analysis in radiological ROC studies.** *Invest Radiol* 1989; 24:234–245
7. Hanley JA, McNeil BJ. **The meaning and use of the area under a receiver operating characteristic (ROC) curve.** *Radiology* 1982;143:29–36
8. Krouwer HGJ, Kim TA, Rand SD, et al. **Single-voxel proton MR spectroscopy of nonneoplastic brain lesions suggestive of a neoplasm.** *AJNR Am J Neuroradiol* 1998;19:1695–1703
9. Fulham MJ, Bizzi A, Dietz MJ, et al. **Mapping of brain tumor metabolites with proton MR spectroscopic imaging: Clinical relevance.** *Radiology* 1992;185:675–686
10. Hollis DP. ***Abusing Cancer Science: The Truth about NMR and Cancer.*** Chehalis: The Strawberry Fields Press;1987