Decision Memo for Magnetic Resonance Spectroscopy for Brain Tumors (CAG-00141N)

Decision Summary

The Centers for Medicare and Medicaid Services (CMS) has determined that the evidence is not adequate to conclude that Magnetic Resonance Spectroscopy (MRS) is reasonable and necessary for diagnosis of brain tumors; therefore, we will continue the current national noncoverage determination.

Back to Top

Decision Memo

This decision memorandum does not constitute a national coverage determination (NCD). It states the intent of the Centers for Medicare & Medicaid Services (CMS) to issue an NCD. Prior to any new or modified policy taking effect, CMS must first issue a manual instruction, program memorandum, CMS ruling or Federal Register Notice, giving specific directions to our claims processing contractors. That issuance, which includes an effective date, is the NCD. If appropriate, the Agency must also change billing and claims processing systems and issue related instructions to allow for payment. The NCD will be published in the Medicare Coverage Issues Manual. Policy changes become effective as of the date listed in the transmittal that announces the Coverage Issues Manual revision.

To: Administrative File: CAG 00141
Magnetic Resonance Spectroscopy for Brain Tumors

From:

Steve E. Phurrough, MD MPA
Director, Coverage and Analysis Group

Jesse M. Polansky, MD, MPH
Director, Division of Items and Devices

Stuart Caplan, RN, MAS
Lead Analyst

Shamiram R. Feinglass, MD, MPH
Lead Medical Officer

Carlos Cano, MD
Medical Officer

Subject: Magnetic Resonance Spectroscopy for Brain Tumors
Date: January 29, 2004

I. Decision

The Centers for Medicare and Medicaid Services (CMS) has determined that the evidence is not adequate to conclude that Magnetic Resonance Spectroscopy (MRS) is reasonable and necessary for diagnosis of brain tumors; therefore, we will continue the current national noncoverage determination.

II. Background

Magnetic resonance spectroscopy (MRS) is a non-invasive diagnostic test that uses strong magnetic fields to measure and analyze the chemical composition of human tissues. MRS relies on the fact that chemicals in the body emit radiofrequency signals when stimulated by a strong magnetic field. By analyzing the different chemical compounds or metabolites in a diseased tissue area (e.g., in the brain) and comparing these with the normal metabolite composition of corresponding tissue, MRS has the potential to provide information that can assist in diagnosing pathologic states.

For the diagnosis of indeterminate brain lesions, MRS is usually performed as an adjunct to a conventional magnetic resonance imaging (MRI) device equipped with the appropriate additional MRS software sequences. MRI images are first obtained to identify the region of interest and guide the procedure. The localized block of tissue to be studied is typically 1-2 cm in each dimension and is known as a voxel. (Voxel size, and the number of voxels studied, may be modified from test to test.) The MRS software sequences decode the radiofrequency signals the strong magnetic field elicits from the tissue of interest and generates a wave-like graph containing peaks that correspond to the various chemical components or metabolites present. Each metabolite has a characteristic radiofrequency signal and a specific location along the horizontal axis of the graph. The height of the various peaks reflects the concentration (signal strength) of the respective chemical component in the tissue under study.

The distribution of peaks and their height define a "metabolite profile" or "spectrum" for the voxel in question. Ratios comparing the concentration of any two chemical compounds in a given area of the brain can be similarly generated. Thus, using MRS to develop "normal" spectra and ratios for various tissues and then comparing these with "abnormal" spectra or ratios obtained from patients with brain lesions could potentially assist in the diagnosis of indeterminate lesions and the care management of these patients. (An abnormal spectrum obtained from a lesion can be compared with that of a control region from the same patient or those from controls matched for age and other variables.)

Diagnosing and treating space-occupying tumors of the brain presents special challenges. This is due in part to the similar appearance of brain tumors and other pathologic entities on computer tomography (CT) or MRI, and the similar appearance as well of brain tumor cell types on such conventional neuroimaging. The inaccessibility of these lesions and their proximity to critical brain structures also complicates their diagnosis. A non-invasive technique that could provide information about the chemical and histologic composition of brain tissue could aid in the diagnosis and treatment of brain tumors by helping to avoid unnecessary biopsies, by guiding biopsies, and by providing additional information for improving treatment.

The most commonly performed MRS technique uses hydrogen nuclei (protons) to obtain a spectrum. Although MRS has been performed using other nuclei (e.g., nitrogen, phosphorus, or carbon,) proton MRS has been the primary focus of previous assessments for this diagnostic test. CMS has also centered this review on proton MRS. In normal brain tissue, proton MRS metabolite profiles typically contain peaks of N-acetylaspartate (NAA,) as well as choline (Cho) and creatine (Cr). In abnormal spectra, lactate (Lac) and lipid (Lip) are also detectable, and the level of choline can also be elevated. For instance, MRS findings characteristically associated with non-necrotic tumors include elevated Cho levels and reduced NAA levels. The presence of high lactate peaks has been reported as the result of anaerobic glycolysis in glioblastoma and brain abscess. The most frequently studied chemical ratios to distinguish tumors from other brain lesions with MRS have been Cho/Cr, Cho/NAA, and Lac/Cr. Specifically, a Cho/NAA ratio of greater than 1 is generally considered to be positive for neoplasm.[1]

MRS has been evaluated as a diagnostic tool for a variety of diagnostic applications. In this decision memorandum, we exclusively review the use of MRS in distinguishing indeterminate lesions in the brain. We also look at the proposed use of MRS as an aid in conducting brain biopsies.

III. History of Medicare Coverage

CMS has determined that MRS falls under the following benefit categories in accordance with the Social Security Act:

- Section 1861(b)(3), inpatient diagnostic services
- Section 1861(s)(1), physicians' services
- Section 1861(s)(3), outpatient diagnostic services

On March 22, 1994, CMS considered MRS an investigational procedure and issued a national noncoverage determination. Accordingly, all uses of MRS are currently noncovered nationally.[2] As indicated in the timeline below, on August 8, 2002, CMS accepted a request to reconsider this noncoverage policy and began a national coverage determination process to review the evidence available on the use of magnetic resonance spectroscopy (MRS) for brain tumors.

IV. Timeline of Recent Activities

August 8, 2002	The American College of Radiology (ACR) formally requested that CMS reconsider its noncoverage determination concerning MRS for the following indications: 1. Cerebral tumor versus abscess or other infectious or inflammatory process, and 2. Cerebral tumor versus radiation necrosis.
September 5, 2002	CMS accepted the formal request for a national coverage determination, and opened the issue for public comment.

30-day public comment period:	The 30-day comment period extends from September 5, 2002 through October 7, 2002.
November 5, 2002	CMS requested that the Agency for Healthcare Research and Quality (AHRQ) commission a technology assessment (TA) of magnetic resonance spectroscopy for brain tumors.
February 10, 2003	TA questions finalized and posted on the Medicare coverage website.
June 17, 2003	TA received and posted on coverage website.

V. FDA Status

The Food and Drug Administration (FDA) has cleared magnetic resonance devices, along with various software packages used to perform proton spectroscopy (MRS), for general diagnostic use through the 510(k) clearance process.

As we stated in 66 FR 58788, 58797 (November 23, 2001), "[t]he criteria the FDA uses in making determinations related to substantial equivalency under section 510(k) of the Food, Drug, and Cosmetic Act is significantly different from the scientific evidence we consider in making "reasonable and necessary" determinations under Medicare. FDA does not necessarily require clinical data or outcomes studies in making a determination of substantial equivalency for the purpose of device approval under section 510(k) of the Food, Drug, and Cosmetic Act. Medicare NCDs consider medical benefit and clinical utility of an item or service in determining whether the item or service is considered reasonable and necessary under the Medicare program. Thus, a substantial equivalency approval under section 510(k) of FDA is not sufficient for making a determination concerning Medicare coverage."

VI. General Methodological Principles

When making national coverage determinations, CMS evaluates relevant clinical evidence to determine whether or not the evidence is of sufficient quality to support a finding that an item or service is reasonable and necessary. The overall objective for the critical appraisal of the evidence is to determine to what degree we are confident that: 1) the specific assessment questions can be answered conclusively; and 2) the intervention will improve net health outcomes for patients.

We divide the assessment of clinical evidence into three stages: 1) the quality of the individual studies; 2) the generalizability of findings from individual studies to the Medicare population; and 3) overarching conclusions that can be drawn from the body of the evidence on the direction and magnitude of the intervention's potential risks and benefits.

The methodological principles described below represent a broad discussion of the issues we consider when reviewing clinical evidence. However, it should be noted that each coverage determination has its unique methodological aspects.

A. Assessing Individual Studies

Methodologists have developed criteria to determine weaknesses and strengths of clinical research. Strength of evidence generally refers to: 1) the scientific validity underlying study findings regarding causal relationships between health care interventions and health outcomes; and 2) the reduction of bias. In general, some of the methodological attributes associated with stronger evidence include those listed below:

- Use of randomization (allocation of patients to either intervention or control group) in order to minimize bias.
- Use of contemporaneous control groups (rather than historical controls) in order to ensure comparability between the intervention and control groups.
- Prospective (rather than retrospective) studies to ensure a more thorough and systematical assessment of factors related to outcomes.
- Larger sample sizes in studies to demonstrate both statistically significant as well as clinically significant outcomes that can be extrapolated to the Medicare population. Sample size should be large enough to make chance an unlikely explanation for what was found.
- Masking (blinding) to ensure patients and investigators do not know to which group patients were assigned (intervention or control). This is important especially in subjective outcomes, such as pain or quality of life, where enthusiasm and psychological factors may lead to an improved perceived outcome by either the patient or assessor.

Regardless of whether the design of a study is a randomized controlled trial, a non-randomized controlled trial, a cohort study or a case-control study, the primary criterion for methodological strength or quality is the extent to which differences between intervention and control groups can be attributed to the intervention studied. This is known as internal validity. Various types of bias can undermine internal validity. These include:

- Different characteristics between patients participating and those theoretically eligible for study but not participating (selection bias).
- Co-interventions or provision of care apart from the intervention under evaluation (performance bias).
- Differential assessment of outcome (detection bias).
- Occurrence and reporting of patients who do not complete the study (attrition bias).

In principle, rankings of research design have been based on the ability of each study design category to minimize these biases. A randomized controlled trial minimizes systematic bias (in theory) by selecting a sample of participants from a particular population and allocating them randomly to the intervention and control groups. Thus, in general, randomized controlled studies have been typically assigned the greatest strength, followed by non-randomized clinical trials and controlled observational studies. The design, conduct and analysis of trials are important factors as well. For example, a well designed and conducted observational study with a large sample size may provide stronger evidence than a poorly designed and conducted randomized controlled trial with a small sample size. The following is a representative list of study designs (some of which have alternative names) ranked from most to least methodologically rigorous in their potential ability to minimize systematic bias:

- Randomized controlled trials
- Non-randomized controlled trials
- Prospective cohort studies
- Retrospective case control studies
- Cross-sectional studies
- Surveillance studies (e.g., using registries or surveys)
- Consecutive case series
- Single case reports

When there are merely associations but not causal relationships between a study's variables and outcomes, it is important not to draw causal inferences. Confounding refers to independent variables that systematically vary with the causal variable. This distorts measurement of the outcome of interest because its effect size is mixed with the effects of other extraneous factors. For observational, and in some cases randomized controlled trials, the method in which confounding factors are handled (either through stratification or appropriate statistical modeling) are of particular concern. For example, in order to interpret and generalize conclusions to our population of Medicare patients, it may be necessary for studies to match or stratify their intervention and control groups by patient age or co-morbidities.

Methodological strength is, therefore, a multidimensional concept that relates to the design, implementation, and analysis of a clinical study. In addition, thorough documentation of the conduct of the research, particularly study selection criteria, rate of attrition and process for data collection, is essential for CMS to adequately assess and consider the evidence.

B. Generalizability of Clinical Evidence to the Medicare Population

The applicability of the results of a study to other populations, settings, treatment regimens and outcomes assessed is known as external validity. Even well-designed and well-conducted trials may not supply the evidence needed if the results of a study are not applicable to the Medicare population. Evidence that provides accurate information about a population or setting not well represented in the Medicare program would be considered but would suffer from limited generalizability.

The extent to which the results of a trial are applicable to other circumstances is often a matter of judgment that depends on specific study characteristics, primarily the patient population studied (age, sex, severity of disease and presence of co-morbidities) and the care setting (primary to tertiary level of care, as well as the experience and specialization of the care provider). Additional relevant variables are treatment regimens (dosage, timing and route of administration), co-interventions or concomitant therapies, and type of outcome and length of follow-up.

The level of care and the experience of the providers in the study are other crucial elements in assessing a study's external validity. Trial participants in an academic medical center may receive more or different attention than is typically available in non-tertiary settings. For example, an investigator's lengthy and detailed explanations of the potential benefits of the intervention and/or the use of new equipment provided to the academic center by the study sponsor may raise doubts about the applicability of study findings to community practice.

Given the evidence available in the research literature, some degree of generalization about an intervention's potential benefits and harms is invariably required in making coverage determinations for the Medicare population. Conditions that assist us in making reasonable generalizations are biologic plausibility, similarities between the populations studied and Medicare patients (age, sex, ethnicity and clinical presentation) and similarities of the intervention studied to those that would be routinely available in community practice.

A study's selected outcomes are an important consideration in generalizing available clinical evidence to Medicare coverage determinations. The goal of our determination process is to assess net health outcomes. These outcomes include resultant risks and benefits such as increased or decreased morbidity and mortality. In order to make this determination, it is often necessary to evaluate whether the strength of the evidence is adequate to draw conclusions about the direction and magnitude of each individual outcome relevant to the intervention under study. In addition, it is important that an intervention's benefits are clinically significant and durable, rather than marginal or short-lived.

If key health outcomes have not been studied or the direction of clinical effect is inconclusive, we may also evaluate the strength and adequacy of indirect evidence linking intermediate or surrogate outcomes to our outcomes of interest.

C. Assessing the Relative Magnitude of Risks and Benefits

An intervention is not reasonable and necessary if the overall risk of harm to the Medicare population is substantial in relation to health benefits. For all determinations, CMS evaluates whether reported benefits translate into improved net health outcomes. CMS places greater emphasis on health outcomes actually experienced by patients, such as quality of life, functional status, duration of disability, morbidity and mortality, and less emphasis on outcomes that patients do not directly experience, such as intermediate outcomes, surrogate outcomes, and laboratory or radiographic responses. The direction, magnitude, and consistency of the risks and benefits across studies are also important considerations. Based on the analysis of the strength of the evidence, CMS assesses the relative magnitude of an intervention or technology's benefits and risk of harm to Medicare beneficiaries.

D. Assessing Diagnostic tests

While the critical appraisal of a diagnostic test follows the general principles described above, there are additional considerations. In making coverage determinations, CMS staff assesses whether valid study results indicate that the test is accurate enough to distinguish patients with and without a target disorder. Second, we must assess if the test is applicable to the Medicare population and likely to change patient management and improve final health outcomes. The following questions address the validity of evidence about the accuracy of diagnostic tests.[3]

- Was there an independent, blind comparison with a reference standard of diagnosis?

The evidence should meet two criteria. First, the patients in the study should have undergone both the diagnostic test under study and the reference standard (e.g., an autopsy or biopsy "proving" that they do or do not have the disorder). Second, those who are applying and interpreting the results of one test should not know the results of the other (e.g., the pathologist interpreting the biopsy constituting the reference standard should be "blind" to the result of the test under study). Thus, investigators can avoid the bias that might result, from example, from "overinterpreting" the reference standard test when the one under study is positive or "underinterpreting" it when negative. For instance, histologic diagnosis and clinical follow-up have been used as reference standards for MRS.

It is key that investigators apply the reference standard regardless of the diagnostic test result. When the reference standard is invasive or risky, an alternative reference standard might be ensuring that the patient does not suffer any adverse outcomes during a long follow up in the absence of treatment and thus likely did not have the adverse disorder when the test was performed.

- Was the diagnostic test evaluated in a range of patients similar to those in whom the test would be used in practice?

Useful articles apply the test under study to patients with mild and severe disease, early and late cases of the target disorder, and among treated and untreated patients. In particular, diagnostic tests should be applied to individuals with different disorders that are commonly confused with the target disorder of interest, rather than, for example, florid cases vs. asymptomatic volunteers.

Valid study results make possible comparisons between the accuracy of the diagnostic test under study and other diagnostic modalities. Accuracy refers to the ability of the test to distinguish patients who have or do not have the target disorder. Measures used to determine accuracy include sensitivity (probability of a positive test result in a patient with the disease) and specificity (probability of a negative test in a patient who does not have the disease). Rather than a combined measure of accuracy, sensitivity and specificity should be separately reported in a study.

Even though a diagnostic test may be accurate if the information it provides does not alter the patient's management, CMS may determine that the test is not reasonable and necessary. In general, diagnostic tests likely to affect patient management are those that provide information that produce large changes between the pre-test and post-test probability that a patient may have the target disorder.

VII. Evidence

A. Introduction.

Consistent findings across studies of net health outcomes associated with an intervention or diagnostic test as well as the magnitude of its risks and benefits are key to the coverage decision process. For this decision memorandum, CMS commissioned a TA from AHRQ that reviewed the published clinical evidence on MRS to determine if MRS might aid in the diagnosis and treatment of brain tumors and improve patient outcomes (in particular by diminishing unnecessary biopsies.) In addition to reviewing the commissioned TA, CMS staff evaluated the individual clinical studies included in that document, and also reviewed another recently published TA report produced by the Technology Evaluation Center (TEC) of the Blue Cross and Blue Shield Association (BCBSA).

As indicated above, MRS has generally been proposed as an adjunct (rather than a substitute) for MRI and CT. In that setting, any added diagnostic accuracy could potentially result in a reduction in biopsy procedures for a proportion of previously indeterminate brain lesions, an improved net health outcome for this population if the biopsies forgone were not necessary. Accuracy (i.e., sensitivity and specificity) of MRS constitutes an intermediate outcome in this context, and histological sampling (i.e., biopsy, culture) must be the reference standard to measure test performance characteristics such as sensitivity and specificity.

B. Discussion of evidence reviewed

1. Analytic questions

The development of an assessment in support of Medicare coverage decisions is based on the same general question for almost all requests: "Is the evidence sufficient to conclude that the application of the technology under study will improve net health outcomes for Medicare patients?" The formulation of specific questions for the assessment recognizes that the effect of an intervention can depend substantially on how it is delivered, to whom it is applied, the alternatives with which it is being compared, and the deliver setting. In order to appraise the net health outcomes of MRS in comparison with conventional imaging such as MRI or CT, or as an adjunct to these neuroimaging tests, and to identify any relevant patient or operator selection criteria, CMS sought to address the following questions for patients presenting with signs or symptoms of a space-occupying brain lesion:

- For what metabolic profiles does MRS provide equivalent, complementary, or more accurate diagnostic information for (i) initial diagnosis, (ii) recurrence, or (iii) assessing therapy, than the following diagnostic tests?
    - Brain biopsy
    - Conventional anatomic imaging studies
    - MRS plus conventional anatomic imaging studies versus brain biopsy
- Does the use of MRS lead to an improved net health outcome by:
    - Avoiding unnecessary biopsy
    - Obtaining appropriate biopsy, from appropriate location
    - Directing biopsy to an appropriate location
    - Receiving appropriate treatment
    - Avoiding an inappropriate treatment
- Are voxel positions and operator error important factors in obtaining diagnostic images? If so, how do they affect MRS accuracy?

2. External systematic reviews/technology assessments

Systematic reviews are based on a comprehensive and unbiased search of published studies to answer a clearly defined and specific set of clinical questions such as those related to the effectiveness of MRS. A well-defined strategy or protocol (established before the results of the individual studies are known) guides this literature search. Thus, the process of identifying studies for potential inclusion and the sources for finding such articles is explicitly documented at the start of the review. Finally, systematic reviews provide a detailed assessment of the studies included.[4]

In this section, we summarize the findings of two TA reports recently published by the BCBSA TEC and by AHRQ that include systematic reviews on the use of MRS in the differential diagnosis and management of brain tumors.

a) BCBSA TA on MRS for Brain Tumors [5]

Using evidence available through May 2003, this TA addressed whether MRS improves the management and health outcomes in patients being evaluated for suspected brain tumors. The TA identified 2 main groups of patients. The first included those with indeterminate brain lesions that require diagnosis of malignant or non-malignant disease. The second group included patients previously treated for a malignant brain tumor in which post-treatment MRS might help to distinguish between recurrent tumor, necrosis, or other pathologic process.

The TA contained 2 specific assessment questions:

- Does the available evidence demonstrate the sensitivity and specificity of MRS for differentiating neoplastic from non-neoplastic lesions?
- Does the available evidence demonstrate whether MRS improves net health outcomes when used to differentiate neoplastic from non-neoplastic lesions?

A total of 7 studies met the eligibility criteria for inclusion in the TA report. Studies were included if the sample size greater than 10, a method to confirm MRS diagnosis was identified, positive test criteria were specified, and the published data allowed calculation of diagnostic test performance.

Results

Differentiating between recurrent or residual tumor vs. delayed radiation necrosis: One study evaluated 12 pediatric patients previously treated with radiation therapy in whom MRI suggested either tumor recurrence or radiation necrosis. MRS studies in these patients compared choline and creatine profiles. Based on histologic confirmation, MRS sensitivity for identifying recurrent tumor was 71%. Omitting an inconclusive result, the specificity for identifying radiation necrosis was 80%

Differentiating between brain tumor and other non-tumor diagnosis: Five studies evaluated patients with a total of 205 lesions, including known primary tumors and unknown new masses. These studied showed MRS sensitivity values ranging from 79 to 100%, with specificity ranging from 74 to 100%.

Differentiation of intracranial cystic lesions: One study evaluated MRS in patients with intracranial cystic lesions. Although a correct diagnosis was made in 47 of 51 patients, the TA found that MRS interpretation was based on investigator judgment of qualitative metabolite profiles rather than formal criteria to define a positive test.

Effect of MRS on avoiding brain biopsy: Two studies reported on the impact of MRS on patient management. In the first of these studies (n=78) MRS was considered to have a potentially positive influence ion treatment decisions by correctly avoiding biopsy in 29% of cases whereas in 3% of cases, MRS suggested neoplasm and a non-neoplastic condition was found at biopsy or surgery. The other study reported that MRS results were used to avoid biopsy in 7 or 15 cases (46%).

Appraisal

Diagnostic test performance. The TA authors discussed limitations and differences in methods and patient samples among these studies that made it difficult to assess the validity of sensitivity and specificity values for MRS in the clinical situations of interest. Most notably, the studies used different criteria for interpreting positive results. Some investigators evaluated Cho/NAA ratios, while others relied on Cho/Cr ratios. Still others relied on multiple metabolite peaks thresholds to diagnose intracranial masses. Approaches for obtaining MRS spectra also varied, from low to high-strength magnetic fields.

In addition, very small sample size, retrospective or unspecified methods to assemble study sample, variable or inconsistent methods to confirm diagnosis, as well as heterogeneity of patients, were common characteristics in the studies review, resulting in a wide range of sensitivity (79 to 100%) and specificity (74 to 100%) values reported.

Health outcomes. The weaknesses in this body of literature, notably the fact that studies reporting sensitivity and specificity of MRS combined different clinical presentations and different indications in the analysis made it difficult to evaluate the potential utility of MRS for avoiding biopsy.

Citing these limitations of the available evidence, the BCBS report found the scientific evidence inadequate to permit conclusions concerning the effect of MRS on health outcomes.

b) AHRQ TA Report on MRS for Brain Tumors[6]

As mentioned above, CMS commissioned a TA from the Agency for Healthcare Research and Quality (AHRQ) to assess the value of MRS for diagnostic evaluation, surgical planning, and patient management of space-occupying brain tumors. Also requested was a review of factors that may affect the performance of MRS.[7]

Using evidence available through October 4, 2002, this technology assessment concentrated on whether MRS improves net health outcomes in patients being evaluated for brain tumors. The TA identified 96 articles that met inclusion criteria, with 11 of these providing data where sensitivity and specificity could be calculated or where data provided could affect patient management.

Search strategy

An OVID search of the MEDLINE® database was conducted on November 6, 2002. Filters and limitations were used, and inclusion and exclusion criteria developed to identify articles to be reviewed. The search used applicable MeSH headings and textwords and resulted in 959 citations for download and screening. Review of the abstracts resulted in accepting 137 citations that met the criteria for complete article retrieval. In addition, abstracts from recent relevant professional society proceedings were reviewed and included in the analyses.[8]

Ninety-six articles met inclusion criteria for further evaluation. Of these, 85 provided information about technical feasibility only.[9] Studies that consisted predominantly of pediatric patients were excluded from this review. Eleven of the 96 articles provided information beyond the level of technical feasibility to address the assessment questions proposed by CMS.

Results

Of the 11 studies providing data beyond that of technical feasibility, 8 reported on metabolite profiles and the diagnostic information these profiles provided compared to other diagnostic tests. Three small studies reported on the potential impact of MRS on patient management. No study reported on operator error or voxel position in determining the accuracy of diagnostic images. A detailed discussion of methods and results of the studies reviewed can be found below under Internal Assessment (section VII.3)

Appraisal

Diagnostic test performance. The report indicates that some investigators relied on multiple metabolic peaks to diagnose brain lesions. Other investigators emphasized NAA/Cr spectral ratios for the same purpose. Still others emphasized alternative metabolite ratios (e.g., Cho/Cr). The AHRQ TA authors noted that this lack of uniformity in the criteria to define the outcome of interest (e.g., a positive test for a neoplastic lesion) as well as other methodological study weaknesses precluded a thorough evaluation of the diagnostic test performance of MRS for brain tumors.

Net health outcomes. Data acquisition and interpretation differed extensively among the studies reviewed. Citing small sample size and other methodological limitations, the authors of the AHRQ TA report considered that the evidence was inadequate to draw conclusions about the effects of MRS on net health outcomes of patients with brain tumors.

With respect to the overall body of evidence reviewed on MRS, the report concludes that human studies conducted on the use of MRS for brain tumors have demonstrated that this non-invasive method is technically feasible, and have suggested potential benefits for some of the proposed indications. However, standardized techniques for acquiring and interpreting MRS spectra are lacking, and there is a paucity of high quality direct evidence demonstrating the effect of MRS on diagnostic thinking and therapeutic decision-making.

The TA report notes that factors such as the relative rarity of brain tumors, the relatively low installed base of MRS software, and the constraints of clinical practice may have prevented the conduct of large, double-blinded controlled trials that would go beyond exploring technical feasibility of MRS. Experience with MRS has only become available to the general community of radiologists within the past five years. Prior to this time, commercial software sequences for generating and analyzing spectra were not reliable, except in the hands of trained specialists. The current commercial software is vastly improved and can be mastered with a reasonable amount of additional training. Prior to about 1995, MRS was available at only a few research-oriented institutions and studies have typically been single-institution feasibility studies or small case series. MRS is still not available in many community hospitals, and even some academic centers. The recent change in the availability of MRS is only now reaching enough centers to allow more advanced investigations using the technique.

Finally, the TA authors point out that it was initially hoped that tumors would have a characteristic "signature" that would allow prompt and accurate diagnosis of indeterminate brain tumors with MRS. However, such "signatures" have not been found.

3. Internal Technology Assessment

CMS conducted a literature search and did not identify any additional articles over those reviewed in the TA. Below is a detailed summary of the 11 studies identified in the TA report produced by AHRQ, which provided information beyond the technical feasibility of the test relevant to the analytic questions posed by CMS. The information summarized in this section served as the basis for CMS conclusions about the overall adequacy of the evidence in determining whether MRS should be considered reasonable and necessary. Studies are grouped for discussion under the relevant assessment questions. Studies provided data on diagnostic accuracy, diagnostic thinking, and patient management. The articles examined the impact of MRS on either differentiating brain tumors from non-tumors, grading of tumors, differentiating intracranial cystic lesions, or assessing the incremental value of MRS when added to MRI. No articles reported effects on patient net health outcomes

- ***For what metabolite profiles does MRS provide equivalent, complementary, or more accurate diagnostic information than standard diagnostic tests?***

The studies providing data on test performance were further grouped into studies with the main purpose of differentiating tumors from non-tumors (four), grading of tumors (two), differentiating intracranial cystic lesions (one), and assessing the incremental value of MRS added to MRI (one). The purposes of the studies were sufficiently different so that combining or comparing studies within the same group was infeasible.

Results

Differentiating neoplasms from non-neoplasms.

Four studies that met the inclusion criteria addressed this diagnostic outcome. Rand et al. (1997) retrospectively evaluated 55 brain lesions in a consecutive series of 53 patients. The study subject sample was heterogeneous and included patients who had suspected brain neoplasm and patients with a history of previously treated brain tumor with new lesions reflecting either recurrent neoplasia or radiation necrosis. Diagnosis was confirmed by histology in 50/55 cases. The purpose of the study was to measure the accuracy of single-voxel proton MRS in distinguishing normal from abnormal brain tissue and neoplastic from non-neoplastic brain disease. 76% of lesions were brain tumors. In a number of unidentified cases, more than one MRS was performed per patient or lesion.

The MRS spectra were interpreted by one of the four neuroradiologists and one MR spectroscopist who had access to available clinical data and imaging studies. These experienced, unblinded readers interpreted the spectra as diagnostic or not and if diagnostic, as neoplasia or non-neoplasia. The criteria for a positive test was based on investigators interpretation of choline, NAA, and other metabolite levels relative to creatine. Four neuroradiologists blinded to the clinical data and MRI results interpreted MRS spectra retrospectively. Diagnoses were confirmed by histology (50 cases) or clinical follow-up (5 cases).

The sensitivity and specificity of MRS to distinguish between neoplastic and non-neoplastic spectra for the unblinded readers were 95% and 100% respectively. The four blinded readers accumulated 12 false-positive interpretations on eight spectra and 22 false-negative interpretations on 13 spectra. The sensitivity and specificity of MRS to distinguish between neoplastic and non-neoplastic spectra for the four blinded readers averaged 85% and 74%, respectively

Butzen et al. (2000) utilized data from the patient population studied by Rand et al. in 1997. The authors noted that "the most accurate method of clinical MRS interpretation remains an open question." The purpose of this retrospective study was to develop a method to improve the discrimination of neoplasm from non-neoplasm relative to the qualitative interpretation of metabolite spectra or the quantitative interpretation utilizing a Cho/NAA ratio threshold. Ninety-nine consecutive patient spectra (number of patients was not reported) with suspected brain neoplasms or suspected recurrent neoplasia referred for MRS were interpreted by blinded and unblinded readers in a manner similar to the method describe in the previous study. The evaluation with the logistic regression model yielded a sensitivity of 87% and a specificity of 85%. Using a threshold of greater than one for the metabolite ratio Cho/NAA to classify tumors resulted in a 79% sensitivity and a 77% specificity.

McKnight et al. (2002) tested the accuracy of a statistically derived index the authors developed based on Cho/NAA ratios to discriminate neoplastic from non-neoplastic brain lesions. The authors hypothesized that MRS can improve tumor detection when added to MRI imaging with the use of this choline-NAA index (CNI). The CNI score represents the difference between relative Cho and NAA levels in a specific voxel and control voxels for each patient. Sixty-eight patients with suspected gliomas underwent MRI and MRS examinations prior to image-guided resection or biopsy of the tumor. Forty-four of these patients consented to undergo additional spectroscopy-guided biopsies to test the sensitivity and specificity of CNI for distinguishing tumor from non-tumorous tissue.[10] One hundred biopsy samples were thus obtained.

The authors sought to estimate the optimal threshold of the CNI and found that the sensitivity of this test was 90% and the specificity was 86% when the CNI cutoff was set at 2.5. These were based on the number of biopsy samples rather than patients. A sub-analysis examined the ability of the index to differentiate tumor grade. There were 12 tumors with heterogeneous histological findings. At least two biopsies with different histological grades were obtained from 7 of these tumors and in three of these cases, the CNIs did not correlate with the histological grade. The authors conclude that the ability of MRS to predict histological grade of infiltrative brain tumors is poor on a case-by-case basis due to the considerable overlap between the levels of metabolite seen in the different grades of tumor.

Kimura et al. (2001) retrospectively evaluated the correlation between MRS single-voxel metabolite patterns and histopathological findings from lesions that showed a ring-like pattern that were obtained with an image-enhancement MRI technique (gadolinium enhanced). The purpose of the study was to identify spectral patterns characteristic of metastatic brain tumor, glioblastoma, radiation necrosis, abscess or cerebral infarct in order to differentiate these enhanced lesions with single-voxel MRS. Forty-five patients with various brain lesions were studied. Results were compared to histology. Three metabolite ratios were calculated and used for analyses. The investigators found that using a Cho/Cr ratio of 2.48 as the criteria for a positive test, resulted in the point of maximum discrimination between differentiating neoplasm from non-neoplasm. Sensitivity and specificity were not presented in the results but were calculated to be 79% and 81% respectively. The method for patient selection was not reported, and the study was limited to ring-like enhanced lesions found on gadolinium-enhanced MRI.

Clinical Utility of MRS added to MRI.

One case series, by Moller-Hartman et al. (2002), evaluated the clinical utility of MRS added to MRI for the differentiation of intracranial neoplastic and non-neoplastic mass lesions. The study population consisted of a consecutive series of 176 patients with focal intracranial mass lesions using MRI and/or CT imaging. Most cases were histologically verified lesions. All patients underwent a single voxel MRS. Two neuroradiologists independently reviewed the combined MRI and MRS results blinded to the final diagnoses. Two other neuroradiologists independently reviewed only the MRI results blinded to the final diagnoses. A diagnosis was classified as “correct” if the reader correctly assigned the case to the type of intracranial mass lesion and the tumor grade. A “no evidence diagnosis” was assigned if the neuroradiologist could not decide between several diagnoses. Of the 176 spectra, conventional MRI alone made 97 (55.1%) correct diagnoses, 27 (15.3%) incorrect diagnoses, and 52 (29.6%) no evidence diagnoses. MRS added to MRI produced 124 (70.5%) correct diagnoses, 16 (9.1%) incorrect diagnoses, 24 (13.6%) no evidence diagnoses, and 12 (6.8%) examinations without diagnostic value. There was no case in which a correct diagnosis made by MRI alone was interpreted incorrectly by the combination of MRI and MRS. There was no mention of how discrepancies between readers were resolved and no sensitivity or specificity was reported. In addition, there was no report of how MRS actually would effect biopsy or patient management.

Distinguishing among tumors (tumor grading).

A small study by Roser et al. (1997) prospectively evaluated 35 MRS spectra in 17 patients with only suspected glial brain tumors. The purpose of the study was "to apply the metabolic features found in a previous study of 21 healthy controls vs. humans with gliomas to a new cohort of patients with a suspected glial brain tumor and other healthy volunteers." This study sought to both identify and grade glial brain tumors. Stereotactic biopsy or open surgery was performed within a few days after MRS. Single-voxel MRS spectra were acquired. Using data from the earlier study of 21 healthy controls vs. patients with gliomas, the investigators calculated five study-specific, non-generalizable ratios using 6 metabolite resonance measurements. These five metabolite ratios were used in an analysis to construct a two-dimensional graph. All ten cases of glioblastoma multiforme were in the proximity of the high-grade region defined by the aforementioned data. Four of five astrocytomas grade II were classified as low-grade gliomas, and one was classified as high grade. One of the two astrocytomas grade III was classified as high grade and the other as low grade. In addition, the contralateral normal matter of tumor patients was assigned as normal in six cases and low grade in two cases. This study was limited to glial brain tumors and as such, generalizability is questionable.

Tedeschi et al. (1997) prospectively studied 27 patients with known brain gliomas to test the hypothesis that MRS can help detect malignant degeneration and/or recurrence (progressions). The 27 patients received from two to five MRS studies, with a total of 72 MRS imaging studies performed over 3.5 years. The time intervals of the spectra were not reported and clinical reasons for performing the scans were not reported. Multi-voxel spectra were obtained. The investigators used the percentage changes in the normalized Cho signal intensity between two consecutive studies to categorize patients into stable and progressive groups. They found that all progressive cases could be correctly classified using a Cho signal increase of more than 45% and all stable cases had increases of less than 35%. Thus, using a threshold of 40% Cho signal increase between visits, the sensitivity was 100% and specificity was 100%. In addition to the normalized Cho measurements, the investigators also analyzed normalized NAA, Cr, and Lac, as well as the within-voxel metabolite ratios (NAA/Cho, NAA/Cr, Cho/Cr). Other than the normalized Cho measurement, they found no association of the other measurements with disease progression.

Differentiating Intracranial Cystic Lesions.

Shukla-Dave et al. (2001) prospectively evaluated the accuracy of MRS in the differentiation of intracranial cystic lesions but did not comment on the use of MRS to differentiate between neoplasm and non-neoplasm. Fifty-one patients with intracranial cystic lesions on conventional MRI were studied. Single-voxel MRS was performed. Two investigators masked to the MRI results, except that the lesions were cystic, independently interpreted the MRS spectra. There were no formal criteria presented for MRS interpretation; interpretation was based on investigator opinion. The pre-operative diagnosis was based solely on the MRS results. All patients presumably underwent surgery for the intracranial cystic lesions. The final diagnosis was based on the results of histopathology, aspiration and culture of the contents. Fifty MRS spectra out of 51 were interpretable. The criteria for a positive test varied by type of lesion. Of the 51 cases, MRS correctly diagnosed the pathology of intracranial cystic lesions in 46 of 51 (90%) cases; MRS did not contribute to the diagnosis in three cases (6%) and falsely diagnosed benign lesions as malignant in two cases (4%).

- ***Does The Use Of MRS Lead To An Improved Net Health Outcome?***

Results

Change in guiding biopsy.

One small prospective study qualified for this category. The purpose of the study by Hall et al. (2001) was "to determine the utility of intraoperative MRS for targeting during brain biopsy using a skull-mounted trajectory guide." If successful, this would change diagnostic thinking by directing surgeons to the appropriate location and potentially reduce unnecessary biopsies. A review of stereotactic brain biopsies found a diagnostic yield (proportion of biopsies containing useable diagnostic tissue) of 91% (Hall, 1999). A total of 17 cases suspected of brain tumors were evaluated in a prospective study (Hall 2001). All patients had "turbo spectroscopic imaging (TSI)" (a multi-voxel MRS method) and (for purposes of comparison) 7 patients also had single-voxel spectroscopy. MRS spectra were obtained within an intraoperative MRI/MRS suite. The TSI spectra in general had lower spectroscopic resolution and often contained lipid signals that were not evident on single voxel spectra.

All 17 biopsies guided by MRS yielded diagnostic tissues, but it is unclear whether this would differ from other methods of biopsy targeting because there was not a clear comparator. Three lesions did not demonstrate regions of elevated choline on the TSI images, which were later histologically confirmed to be brain tumors.

Change in diagnostic thinking.

A small study by Lin et al. (1999) prospectively evaluated the utility of single voxel MRS as an alternative or adjunct to brain biopsy in patients with lesions on MRI suggestive of brain tumors. In order to determine whether MRS directly impacted upon and altered clinical decision-making, prior to the MRS examination, a neurosurgeon defined a treatment plan that would be carried out in the absence of a diagnostic MRS study. Subsequently, MRS interpretations were directly incorporated into the clinical decision-making and a treatment plan was determined. Patients were then followed to determine if subsequent treatment and outcomes were in accordance or discordance with the MRS findings. Diagnoses were confirmed by biopsy, surgery, or clinical follow up. Single-voxel MRS was performed. Spectra from 15 patients with mixed indications for MRS were analyzed. The criteria for a positive test were based only on investigator interpretation. For 10 patients with previously documented tumors, MRS was interpreted as consistent with recurrent tumors in seven cases and consistent with radiation necrosis in three cases.

In the absence of MRS, the neurosurgeon would have recommended stereotactic biopsy in eight cases, serial MRI at six week intervals in three cases, repeat craniotomy in three cases, and empiric chemotherapy in one case. MRS was used in place of biopsy in seven cases, and correlated with clinical course in six of these cases. Overall, MRS was found to directly alter clinical management in 12 of 15 patients and provided greater support for clinical management in 14 of 15 patients. The authors state that MRS is not a substitute for histology when planning treatment of brain tumors. Because of the small size and lack of specific criteria for analyzing the MRS results, the generalizability of this study is questionable. The authors agree that a larger, prospective study is needed.

Effect on Therapeutic Management.

Adamson et al., (1998) conducted a retrospective review of 78 medical records to assess the influence of single-voxel MRS findings on the treatment of a heterogeneous mix of patients suspected of having a brain tumor. MRS was classified as having a potential positive influence on treatment if no biopsy was needed before the initiation of treatment. If MRS results did not agree with the subsequent clinical diagnosis (either by histology or clinical follow up), the results were considered to have a potential negative influence on patient treatment. In all other cases, the effect of MRS was presumed to be negligible or indeterminate. A Cho/NAA ratio greater than 1.0 was considered to be positive for neoplasm, and thus the criteria for a positive test. MRS was positive for neoplasm in 49 of the 78 patients. In only eight of these 49 patients, MRS was classified as having a potential positive influence. MRS was classified as having a potential negative influence on patient treatment in two of the 49 patients diagnosed as having neoplasm. MRS was negative for neoplasm in 29 of 78 patients. In 15 of these 29 patients, MRS was classified as having a potential positive influence. MRS had no influence on patient treatment in 37 patients diagnosed with brain tumor as the patients still underwent diagnostic testing following MRS. MRS had no influence in 76% of cases where MRS was suggestive of tumor. This represents 47% of all patients in the study.

No study was conducted to evaluate the overall impact of MRS on final health outcomes.

- ***Are voxel positions and operator error important factors in obtaining diagnostic images? If so, how do they impact MRS accuracy?***

No study explicitly evaluated the impact of voxel position on the accuracy of MRS. No study commented on the potential impact of operator error in placement of the voxel.

3. Professional Society Position Statements and Public Comments

CMS received public comments in support of MRS for the indications submitted by the requestors. Nassau Radiologic Group submitted 14 abstracts of published, full-text articles along with their letter of support. Fox Chase Cancer Center submitted one published, full-text article and 5 abstracts of published, full-text articles. University of North Carolina School of Medicine's Department of Radiology submitted 15 full-text articles. Nineteen additional letters of support were submitted, although without supporting literature. All of the aforementioned full-text articles submitted by the public were identified by the TA's selection criteria process and were included in the systematic review.

VIII. CMS Analysis

National coverage determinations (NCDs) are determinations by the Secretary with respect to whether or not a particular item or service is covered nationally under title XVIII of the Social Security Act § 1869(f)(1)(B). In order to be covered by Medicare, an item or service must fall within one or more benefit categories contained within Part A or Part B, and must not be otherwise excluded from coverage. Moreover, with limited exceptions, the expenses incurred for items or services must be "reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member." § 1862(a)(1)(A).

This section summarizes the agency's evaluation of the evidence available on the diagnostic performance characteristics and the effect of MRS on final health outcomes for Medicare beneficiaries. Following a general discussion, we specifically address the three questions that were submitted to AHRQ to guide the external TA and the related evidence that led to the coverage conclusions.

Only 11 of these 96 articles reviewed provided any information beyond that of the technical feasibility of MRS. Thus, the vast majority of the published literature on MRS for brain masses, and the vast majority of comments we received, did not focus on diagnostic performance and patient management–the primary focus of this NCD. Rather, most published studies at this time are restricted to the technical feasibility of MRS of the brain, which addresses only the ability to produce reliable spectra.

This body of evidence was constituted by both retrospective and prospective studies. There was wide variation in the type of MRS performed (e.g., strength of the magnet field utilized varied substantially among studies.). Single voxel sampling was the predominant methodology, while multiple sampling or a combination of the two approaches was sometimes employed. In some articles, the technique was not reported. The range of voxel volumes was not uniform. Rand et al., 1997; Adamson et al., 1998; and Butzen et al., 2000 used overlapping patient samples while addressing different research issues. This overlap in patient population and time period in the two largest studies is noteworthy given the paucity of data on this subject in general . The articles are included here because they report on different outcomes of interest. Post-hoc subgroup analyses, however, are generally regarded as useful in providing suggestive evidence of a technology rather than confirming research hypotheses.

- ***For what metabolite profiles does MRS provide equivalent, complementary, or more accurate diagnostic information than alternative diagnostic procedures?***

Among all the full-text articles examined, only one provided relatively complete reporting of the metabolite signal intensities and ratios for each type of tumor found in their study population (Moller-Hartman 2002). This was also the only study addressing the incremental diagnostic yield of MRS, showing that MRS added to conventional MRI improved the number of correct diagnoses and reduced the number of incorrect or equivocal diagnoses. However, the article did not report whether the two neuroradiologists read all the images or spectra in the same group or how discrepancies between the readers were resolved, which may limit the validity of the results. In addition, MRS's effect on diagnostic accuracy was reported generally, without specific report of specificity, sensitivity or effect on performance of biopsy or other change in management.

Rand (1997) included a variety of diagnoses and multiple blinded and unblinded readers to interpret the spectra results. The patient population in this study was very heterogeneous with regards to diagnosis, making it difficult to determine the potential utility of MRS for each particular condition. In addition, the qualitative interpretation of the MRS spectra diminishes the ability to compare their results with other studies and, more importantly, to generalize results to other settings.

Butzen et al (2000) included the patient population from the study by Rand et al (1997) to retrospectively develop a logistical regression model in order to improve metabolite profile interpretation. The clinical utility of the sensitivity (85%) and specificity (87%) values obtained is difficult to assess given the heterogeneity of the patient population, and requires further confirmation in a prospective study.

McKnight et al (2002) developed a calculated index and based on this study defined a threshold value to differentiate neoplastic from non-neoplastic cerebral lesions on conventional imaging. Tumors studied were restricted to gliomas. The authors do not describe how patients were enrolled in the study nor do they describe the characteristics of patients who declined consent to undergo additional biopsies. As is the case in other studies reporting on the diagnostic performance measures of MRS, the proposed threshold is the result of post-hoc analysis and could be attributable to chance alone. Further studies utilizing the a priori defined threshold must measure its ability to discriminate tumors from non-tumor when compared to histological findings.

The purpose of the study by Kimura et al. was to retrospectively identify characteristic spectral patterns for metastatic tumors, gliomas and other brain lesions based on the known histology of the lesions. All lesions studied had a ring-like appearance and were obtained by a specific image-enhanced MRI technique. Although sensitivity (79%) and specificity (81%) to distinguish neoplastic from non-neoplastic lesions utilizing the proposed choline/creatine ratio could be obtained, the proposed ratio was the result of post-hoc analysis rather than a prior hypothesis and should be tested in subsequent trials to establish accuracy measures for the test. The authors do not report the method of patient selection and only look at a type of lesions obtained with a specialized technique, two limitations precluding the generalizability of the findings.

Tedeschi et al (1997) also had a small sample size. In addition, this study raised a major methodological concern. Not only did the authors not test their findings in a subsequent study but rather modified retrospectively their threshold for a positive test. By doing so, they altered the sensitivity and specificity in the direction of a more favorable result. What is missing is a subsequent confirmatory trial that uses that threshold to separate out other cases and determines the resulting sensitivity and specificity.

Roser (1997) examined tumor grading. In the prospective validation study, all 17 patients had glial brain tumors. Thus, the results of this study cannot be generalized to populations with a broader spectrum of brain lesions. A much larger number of patients with a broader spectrum of brain lesions are needed to develop the diagnostic criteria and to verify the results. Another study examining tumor grading by Tedeschi (1997) used repeated MRS studies. The studies were not based on a fixed time interval, and the clinical reasons for the repeated studies were not explicitly stated, thus limiting clinical generalizability and potential effect on patient management.

Shukla-Dave (2001) made specific diagnoses between different neoplastic etiologies but did not provide a technique to differentiate between non-neoplastic and neoplastic tissue, a distinction that affects patient management. In addition, the spectral interpretation was not based on formal criteria and as such limits reproducibility. As seen in other studies in this area, the rate of discrepancies and the method of resolution of discrepancies in the interpretation of the spectra results between the two investigators were not reported.

It is difficult to draw any substantive conclusions with respect to the diagnostic accuracy of MRS from the literature reviewed given the variability in use and interpretation of metabolic profiles. The multiple peak intensities and ratios of metabolites reported in the AHRQ TA represent a very heterogeneous mix and reflects a lack of consensus in the literature with respect to the metabolite profiles that are in use. Cho/Cr (choline/creatine) is one metabolite ratio that has been found to be useful in differentiating neoplasm and non-neoplasm in several studies. However, some of the signals and ratios were unique for a particular study. The lack of standardization in the choice and interpretation of metabolic profiles may explain the wide range of sensitivity and specificity values reported. It also limits the generalizability of individual study results. Perhaps more importantly, it could lead to inappropriate variation in practice patterns. No single metabolite or ratio, by itself has been shown to distinguish among neoplasms, among different tumor grades, or between neoplastic and non-neoplastic lesions.

In addition, the literature reviewed is not conclusive regarding the potential improved accuracy of MRS compared to conventional diagnostic imaging.

Does The Use Of MRS Lead To An Improved Net Health Outcome?

Although MRS was addressed in a fairly large number of studies, we found a general paucity of data on patient management and patient outcomes. Only three studies (Hall (2001), Lin (1999), Adamson (1998)) addressed the potential impact of MRS results on diagnostic thinking or therapeutic decision-making. As sample sizes were too small to support statistically valid conclusions (n=17 (Hall 2001) and n=15 (Lin)), we do not believe these studies provide evidence of a sufficient quality to provide a basis to conclude that MRS results in improved diagnosis or health outcomes. The only large study, Adamson (1998) with n=90, was a retrospective analysis of medical records to identify potential opportunities for MRS to influence diagnostic thinking.

This issue of sample size is an important limitation for this analysis. Sample sizes that might be adequate for investigating the effect of the diagnostic intervention in one type of tumor are not necessarily adequate for investigating outcomes for multiple types of tumors in the same study. This applies to distinguishing among tumor grades as well.

Hall (2001) prospectively examined the utility of MRS to target brain biopsy. This was a small study (n=17) and was promising as it showed that intraoperative MRS guided biopsy could be safe, simple, and accurate. However, it is limited by the small sample size and the requirement of an intraoperative MRI/S suite, which is not widely available in practice.

Lin (1999) examined MRS as an alternative or adjunct to brain biopsy. Treatment plans put in place by a neurosurgeon were evaluated for changes following use of MRS. Although MRS treatment plans potentially altered treatment in several cases, this study included only 15 patients with a narrow spectrum of diagnoses for MRS and, as such, represents too small of a sample to make any clear conclusions. With only one neurosurgeon involved in the treatment plan, uncertain sampling with no controls, and with MRS spectra quantified using an unspecified standard, the generalizability of this study is limited.

The purpose of the retrospective Adamson study (1998) was to determine the effect of MRS on treatment decisions. There were several limitations in this study. Fourteen of the 78 patients had incomplete follow-up and the decision to do an MRS was based on previous CT/MRI results where neoplasm was considered the primary candidate in the differential diagnosis, thus biasing the study. Additionally, the small number and mixed nature of the patient sample makes it difficult to draw conclusions from this study and limits its generalizability.

These three studies addressed the potential impact of MRS results on diagnostic thinking or therapeutic decision-making. Conclusions that can be drawn from these studies are severely limited due to the fact that the two prospective studies had only 15 and 17 patients, respectively. The only relatively large study was a retrospective analysis of medical records to identify potential opportunities for MRS to influence diagnostic thinking. However, the follow up information (including 12 subjects in the "MRS no tumor" group) was incomplete in this analysis.

- ***Are Voxel Positions And Operator Error Important Factors In Obtaining Diagnostic Images? If So, How Do They Impact MRS Accuracy?***

No study explicitly evaluated the impact of voxel position on the accuracy of MRS. No study commented on the potential impact of operator error in placement of the voxel.

Conclusion

In sum, our analysis of the body of evidence based on standard methodological principles related to the use of proton MRS to differentiate brain lesions that appear on conventional imaging and its effect of patient management and health outcomes is consistent with that of the two external TA reports recently published on this topic. While there are a large number of studies that confirm the technical feasibility of this test, there are very few published studies that evaluate its diagnostic accuracy and whether it can favorably affect diagnostic thinking and therapeutic choice.

The 11 published articles reviewed on the use of MRS for the differential diagnosis of indeterminate brain lesions showed a number of methodological weaknesses that preclude firm conclusions on the validity and generalizability of their findings. Even the wide range of reported findings such as sensitivity and specificity represent a lack of consistent evidence regarding the accuracy of the test. In addition, differences in the criteria for interpreting metabolite profiles or ratios as a positive test, unspecified methods of patient selection, heterogeneity of conditions studied, small sample size, difference in MRS protocols (including low and high strengths of the magnetic field) are common among the studies reviewed.

These methodological shortcomings undermine the confidence in their results. Notably there is no controlled study comparing conventional diagnostic strategies with MRS alone or as an adjunct to demonstrate the effect of interventions with or without MRS on health outcomes. CMS has thus determined that the evidence is not adequate to conclude that MRS is reasonable and necessary for the diagnosis of brain lesions. Therefore, we intend to issue a continuation of noncoverage determination for MRS.

Appendix

Technology Assessment submitted to AHRQ by the New England Medical Center Evidence-based Practice Center, Harmon S. Jordan, ScD; Robert Bert, MD, PhD; Priscilla Chew, MPH; Bruce Kupelnick, BA; and Joseph Lau, MD, Contract No. 290-02-0022, June 13, 2003. The complete TA, including all references and evidence tables, is available online via a hyperlink on the CMS tracking sheet.

---

1 Adamson AJ, Rand SD, Prost RW, Kim TA, Schultz C, Haughton VM. Focal brain lesions: effect of single-voxel proton MR spectroscopic findings on treatment decisions. *Radiology* 1998;209(1):73-78..

2 Section 50-13 of the Medicare Coverage Issues Manual. Magnetic Resonance Imaging.

3 D Sackett, S Straus, W Richardson, W Rosenberg, B Haynes: *Evidence-based Medicine, How to Practice and Teach EBM.* 2nd edition. Churchill Livingstone. 2000

4 Hulley et al. Designing Clinical Research. 2001.

5 Blue Cross Blue Shield Technology Evaluation Center (TEC). Magnetic resonance spectroscopy for evaluation of suspected brain tumor. TEC Assessments 2003;18(1):1-26. http://www.bcbs.com/tec/vol18/18_01.html

6 Technology Assessment submitted to AHRQ by the New England Medical Center Evidence-based Practice Center, Harmon S. Jordan, ScD; Robert Bert, MD, PhD; Priscilla Chew, MPH; Bruce Kupelnick, BA; and Joseph Lau, MD, June 13, 2003

7 The TA report can be found at http://www.cms.gov/mcd/viewtrackingsheet.asp?id=52

8 See the TA for a complete description of the inclusion and exclusion criteria and search strategy

9 See Evidence Table 1 of the TA for a summary of the 85 studies examining technical feasibility

10 In a number of cases, more than one MRS was performed per patient or lesion

Back to Top