Paul A. Bottomley, PhD

The Trouble with Spectroscopy Papers[1]

Writing a critique and guide for authors of clinical spectroscopy research papers is a likely way of ensuring that one never sees another of one's own papers published in this field. Nevertheless, it is disappointing, though perhaps predictable, that despite its historical foundations in quantitative spectroscopy, the field has its fair share of findings that are not so obviously reconciled. Here is the view of one author, one referee, and one spectroscopy protagonist about what might be expected of a clinical spectroscopy paper. In addition to novelty, the fundamental criteria for acceptance should be that the conclusions are supported by properly and objectively quantified results, and that sufficient experimental detail is provided so that one skilled in the art could reproduce the study and its findings.

Index terms: Editorials • Magnetic resonance (MR), spectroscopy

Radiology 1991; 181:344–350

A decade ago, and a year after magnetic resonance (MR) imaging became practical in the body, a report of a phosphorus-31 MR spectroscopy study of a patient with McArdle syndrome (1) heralded the age of clinical spectroscopy research. It put substance to the hope that the brands of chemistry to which the technology provides unique access might now provide new insight into disease, its clinical diagnosis, and treatment. Like early MR imaging, initial spectroscopy studies were, however, often of limited or heterogeneous patient populations and characterized by poor spatial resolution. Criticisms of the statistical rigor of early imaging studies in establishing clinical efficacy (2), as just or unjust as they might be (3,4), might have been leveled at much of the early clinical spectroscopy work, too, if it were not for the fact that, unlike MR imaging (which appeared to be efficacious anyway), MR spectroscopy has yet to find an everyday efficacious clinical application.

Like an author who at first disagrees with a reviewer's comments, then revises his paper to address them, and later concedes that the resulting compromise is a major improvement over the original, we, as clinical spectroscopy researchers, must now strive for robust, quantitative documentation of our clinical findings if we are to expect them to be taken seriously in a medical environment where requirements for demonstrating efficacy, or even establishing fact, grow ever more sophisticated and onerous.

The problems with the introduction of spectroscopy as a new medical technology differ so fundamentally from those of its sister technology, MR imaging, to which it is inextricably linked, that its failure to materialize clinically with the same speed that MR imaging arrived should be of little surprise (5). Whereas MR imaging provided anatomical pictures in the context of a long history of clinical imaging modalities, from x rays to ultrasound (US) and computed tomography, the biochemical information provided by means of spectroscopy has no real clinical antecedents and we must look to the biochemistry research literature for its interpretation. Clinical MR spectroscopy and spectroscopy papers thus fall into relatively uncharted territory in the field of radiology and in this and other clinical journals.

LOUD AND SOFT CONTROVERSIES

That reproducibility is the key to the successful application of spectroscopy to clinical medicine, just as it is the key to scientific research in general, goes without saying. It is also easier said than done. The Table notes a few examples of different findings from different laboratories, starting with the controversy over the proton (H-1) spectroscopy blood test for cancer (6–21). In general, discrepencies are attributable to different systematic error components in the measurements, which include the use of different MR techniques, timing parameters, patient protocol, and disease status. Because there are often more than one source of systematic error and each error may add or subtract from the measured quantities, a complete objective accounting of the sources of differences is rarely feasible on the basis of what one finds, at least in the bodies of present published articles. The onus is on authors to acknowledge and provide some attempt at a feasible explanation for any differences between their values and previously published quantitative values when they occur.

This is why the exact protocol and methods employed for each study must be meticulously documented. Things left unstated tend to be misstated. In my own case, a seemingly unimportant omission of stating in print between 1982 and 1984 that the time required to switch our MR imaging instrument from H-1 imaging to P-31 spectroscopy was typically less than 90 seconds (22) led to the misrepresentation that the switchover time was 2–14 hours in an argument against the technology (Kaufman L, Crooks L, Margulis H. MRI at the threshold. Diagnostic Imaging 1984;[Jan]:36–43).

---

[1] From the GE Research and Development Center, NMR Bldg, Rm 130, River Rd, PO Box 8, Schenectady, NY 12309. Received July 22, 1991; accepted July 23. **Address reprint requests to** the author.
© RSNA, 1991

Reproducibility in Clinical Applications of MR Spectroscopy

Reference	Claims and Counterclaims	Evidence	Comments and Problems
6	Claim: cancer detection by using plasma lipid H-1 line widths	No overlap in values between patients with malignant tumor and healthy subjects	
7,8	Counterclaim: plasma line widths unreliable for cancer detection	Much overlap between patients and healthy subjects	Problems: Variable plasma triglyceride levels cause excessive false-positive and false-negative detections. Sample treatment
9	Claim: clinical detection of heart transplant rejection by using plasma lipid H-1 line widths	90% predictive value for detecting rejection: normal widths in patients indicate rejection, but abnormal widths mean no rejection	Comment: Can patients with rejection of sufficient severity to require augmented therapy be distinguished from those who do not? Why do normal widths signify rejection and abnormal widths signify no rejection?
10	Claim: correlation between plasma glycosylated residues and heart transplant rejection	Useful correlations between echocardiographic data and MR imaging in only eight of 13 patients	Problem: reliability for predicting rejection on an individual basis
11	Claim: brain T1(PCr) = 4.8 ± 0.2 sec, T1(γ-ATP) = 1.0 sec	Topical MR spectroscopy in subjects at 1.9 T	Comment: differences probably due to systematic error. Absolute values varying this much may affect quantitation if used for correction, depending on repetition time. What are the correct values?
12	Counterclaim: brain T1(PCr) = 3.1 ± 0.5 sec (1.5 T), T1(PCr) = 2.7 ± 0.2 sec (2.0 T), T1(γ-ATP) = 0.6 ± 0.1 sec (1.5–2 T)	ISIS in seven subjects at 1.5 T, three subjects at 2.0 T	
13	Counterclaim: brain T1(PCr) = 3.1 ± 0.5 sec, T1 (γ-ATP) = 1.36 ± 0.3 sec	ISIS at 1.5 T	
14	Claim: normal heart PCr/ATP = 1.6 ± 0.4	DRESS, six subjects	Problems: possible chest muscle contamination (increases PCr/ATP); acquisition delay (decreases PCr/ATP); partial saturation (usually decreases PCr/ATP, especially with ISIS); cardiac motion in ISIS; blood ATP contamination (decreases PCr/ATP)
15	Claim: normal heart PCr/ATP = 1.6 ± 0.2	Rotating frame, six subjects	
16	Claim: normal heart PCr/ATP = 1.72 ± 0.15	One-dimensional phase encoding, 11 subjects	
17	Claim: normal heart PCr/ATP = 1.93 ± 0.21	One-dimensional phase encoding, 19 subjects	
18	Claim: normal heart PCr/ATP = 1.7 ± 0.2	Three-dimensional spectroscopic imaging, four adults	
19	Counterclaim: normal heart PCr/ATP = 1.33 ± 0.19	ISIS, 10 subjects	
20	Counterclaim: normal heart PCr/ATP = 1.37 ± 0.09	ISIS, 10 subjects	
21	Counterclaim: normal heart PCr/ATP = 0.89 ± 0.08	Three-dimensional phase encoding, 14 subjects	

Note.—ATP = adenosine triphosphate, DRESS = depth-resolved surface coil spectroscopy, ISIS = image-selected in vivo spectroscopy, PCr = phosphocreatine.

PICTURES, PEAKS, AND NUMBERS

From the outset, primary findings with spectroscopy were displayed as graphs of the absorption-mode intensity versus chemical shift in parts per million and quantified numerically as metabolite ratios, pH units, or concentrations, not pictures. This mode of quantification suits spectroscopy well to statistical validation techniques and even blinded studies, but at least quantitative numerical comparisons. It is indeed a technology well suited to the types of quantitative analysis embraced for scientifically rigorous clinical research (2).

Reformatting localized spectroscopic information into anatomic-style images whose intensities or hues are proportional to a metabolite or the ratio of two does not of itself, without additional quantification, contribute to the establishment of a robust scientific finding, which might then be found to be of clinical value. Such displays alone do not address the problem of interpretation posed by the unique character of the information, nor do they necessarily facilitate acceptance of the technology in radiology. Their generally poor spatial resolution and vulnerability to artifact seldom render them usable as anatomic images in their own right, which is why their interpretation in conjunction with another imaging modality, such as MR imaging or US, has become a virtual necessity. There is a danger that if spectroscopy is proffered only as an anatomic imaging modality, it will be judged as one. As such, it would almost certainly fail because millimole-per-liter tissue metabolites cannot be imaged with MR with the same quality as 100-mole-per-liter tissue water protons (5).

There is no objection to the inclusion of intensity or color-coded spectroscopic images in a clinical paper as an aid to locating or visualizing certain results (this may well be their forte), but the quantitative analysis of the metabolic information that establishes the basic findings should be a primary consideration. If we are then to change from a numerically based to an image-based quantification in spectroscopy, a basis for reliable and reproducible presentation and interpretation of spectroscopic images must be established; otherwise we will be moving backward.

Thus, if spectroscopy findings are to be evaluated by means of images alone, there must be a proper accounting of all the artifacts and abnormalities manifested in images obtained from patients, as well as from healthy control subjects, throughout the field of view, not just those abnormalities or artifacts that happen to coincide with, say, the location of a known pathologic condition as determined with MR imaging. It must be assumed, until proved otherwise, that the metabolic information at the millimole-per-liter level is independent of the MR imaging information to which adipose and tissue water are the overwhelming contributing moieties. Indeed, many examples already exist in which metabolic abnormalities demonstrated by means of spectroscopy have no obvious correlation with the anatomic characteristics of a pathologic condition as seen with MR imaging: the dementia complex associated with acquired immunodeficiency syndrome (23,24); tumors

(25); and myocardial ischemia (16), to name a few. The standard deviation of the noise in a spectroscopic image should be less than the magnitude of the change being claimed for the metabolite signal intensity in a particular voxel, and this noise should not be obfuscated by careful windowing in "representative" examples. A reading of spectroscopic images by multiple blinded observers would do much to allay these concerns.

NUMBERS FROM PEAKS

What should the numbers measure, or if one must display a spectroscopic image, what should its intensity or hue reflect? If all spectral peaks were nonoverlapping singlets acquired under conditions in which neither the instrument nor the subject contributed significant inhomogeneous broadening to the spectral widths in all data sets constituting a given study, the peak heights might provide an acceptable measure of their source moieties. This is rarely true or achievable in today's clinical spectroscopy examinations, however, and such a claim would be difficult to validate and just as difficult to reproduce by others.

Unlike peak heights, the integrated area of an absorption-mode resonance is relatively immune from instrument- and subject-induced line broadening and from the complication of multiplet line structure. It should therefore form the basis of all measurements of metabolite levels used for interexamination comparisons. Magnitude spectra, calculated from the root of the sum of the squares of the absorption and dispersion modes, contain broader peaks and signal artifacts when any broad baseline components are present, as is often the case in vivo, and are therefore usually unacceptable, except perhaps when they are derived from spin echoes. Peak heights of absorption spectra are suitable for intraexamination comparisons when nothing is changed that could conceivably alter the line widths, which includes moving the patient.

Perversely, the areas of the peaks of most interest most often overlap others, which means that the traditional running integral used by chemical MR spectroscopists is unacceptable, as illustrated in Figure 1. Area measurements of overlapping peaks that are delimited by saddle points are also generally less reliable if either of the adjacent peaks varies independently. This is because the relative location of the saddle point will change as the peaks vary in amplitude, and most of the signal in a peak resides near its base, not its summit. Therefore, the individual peaks must be deconvolved when they overlap.

Such deconvolution typically involves curve fitting a spectrum with Lorentzian, Gaussian, or a Lorentzian-Gaussian mixture of line shapes—whichever fits best—so that the difference between the original and the fitted spectra are less than or equal to the noise. The number used to characterize each component resonance is then the integral of its corresponding fitted curve. Before curve fitting, spectra must usually be conditioned so that they are in fact in the absorption mode, which may involve zero and first-order (varying linear with frequency in parts per million) phase corrections. Any broad baseline signal components must be accounted for, usually by fitting or subtraction, to avoid significant contamination of the peaks of interest with spurious baseline area contributions.

Overlap is especially problematic for the quantification of Pi in P-31 heart spectra, where it falls but a few tenths of 1 ppm from one of the blood 2,3 diphosphoglycerate (DPG) resonances in the PM region, and for the quantification of lactate in H-1 spectra, which falls in a region of the spectrum prone to contain a broad lipid peak. Since lipids are ubiquitous in the body, conclusions about lactate detection in pathologic conditions and the assignment of signals in the lactate region of a H-1 spectrum wholly to lactate require evidence: "It isn't lactate 'til the fat peak is discounted" should be the maxim adopted for quantitative studies. When pH measurements are derived from measurements of the chemical shift of a curve-fitted Pi peak in the presence of overlapping PM in a P-31 spectrum, expect to see baseline, calibration, and sampling uncertainties of about 0.04 to 0.1 pH unit when the signal-to-noise ratio (S/N) is 3 or more, and up to 0.14 pH unit or more with a S/N of 1–3, even for narrow line widths (26).

Spectral processing is automatable (27,28) if appropriate care is taken to ensure that the algorithms are robust to spectra of variable quality and with spurious artifacts. There are also valid spectral processing techniques that can be performed on the raw time-domain data (29,30). These techniques are analogous to those noted above in that they involve curve fitting of the free-induction-decay signal with a decaying function such as the inverse Fourier transform of a Lorentzian (that is, an exponential) or a Gaussian function for each resonance. In each case, it is important to obtain an estimate of the noise or the S/N as an index to cross-check reliability. Since processing will vary among studies and sites, details of the conditioning applied should be fully documented if that hypothetically skilled-in-the-art reader is to have a chance at reproducing the findings.

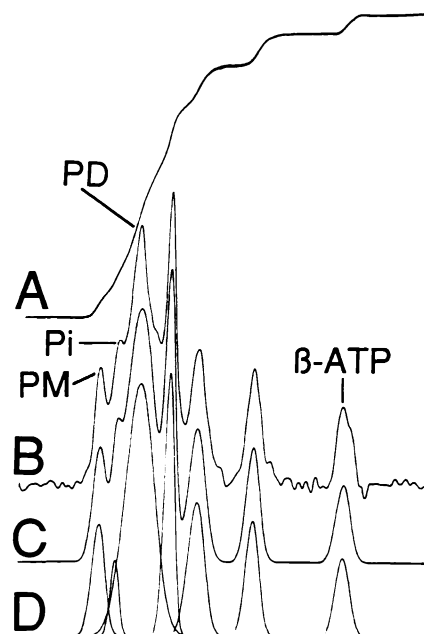
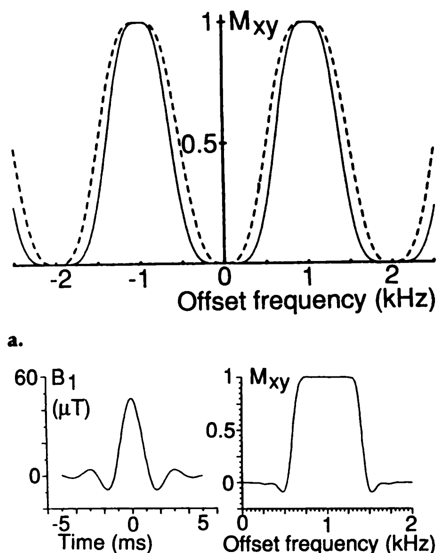


Figure 1. How do you quantify a P-31 brain spectrum? *A* = conventional running line integral. Inflection points are unresolved, but integration of the signal between the saddle points of the respective peaks, including the contaminating phosphodiester (*PD*), yields PCr/β-ATP of 3.4 and PCr/inorganic phosphate (*Pi*) of 2.0. *B* = spectrum from a 3-cm-thick axial section through the brain of a patient with suspected Alzheimer dementia in whom a large pituitary tumor and infarction were seen at MR imaging during the spectroscopy examination (brain area, 135 cm^2; tumor and infarct area, 25 cm^2; acquisition made at 1.5 T with a repetition time of 15 seconds, a one-dimensional ISIS sequence, a uniform P-31 head coil, and a 0.5-msec acquisition delay) (23). Phosphomonoester (*PM*), Pi, and PCr peaks overlap that of PD. By using peak heights, PCr/β-ATP = 3.8, and PCr/Pi = 2.1. Curve fitting the spectrum (*C*) to resolve the resonances (*D*) yields PCr/ATP of 3.4 and PCr/Pi of 3.5 if the heights of the component peaks are used, or PCr/ATP of 2.3 and PCr/Pi of 4.1 by using the integrals of the component peaks.

The widespread use of ratios of the metabolite integrals, as opposed to the reporting of integrated areas in arbitrary units, appropriately accounts for the latter generally varying from study to study with spectrometer gain, voxel size, and other systematic differences. Accordingly, tables of arbitrary peak areas are of little use without normalization. A recent odd practice of normalizing peak areas by dividing them by the total integrated area of the spectrum, however, should be acceptable only if all of that area is fully assigned and accounted for (lest there be changes in unidentified components) and if there are no artifacts present in the total spectrum that vary over the course of a study (lest they systematically or randomly affect the findings).

Measurements of the absolute concentrations of metabolites, rather than their ratios, have become increasingly



b.

Figure 2. Selective excitation bandwidths **(a)** of 1-$\bar{3}$-3-$\bar{1}$ (solid curve) and 1-$\bar{5}$-10-$\overline{10}$-5-$\bar{1}$ (dashed curve) solvent suppression 90° composite pulses with interpulse delays of 0.5 msec (adapted, with permission, from reference 35) and **(b,** at right) of a single 10-msec cosine-windowed 90° chemical-selective sinc pulse (at left) at a 1-kHz offset frequency (36). M_{xy} = transverse nuclear magnetization, B_1 = transverse excitation field amplitude.

important in characterizing which metabolites are and are not altered in various disease states. These measurements are performed by comparing the peak integrals with that of a concentration reference, but an accurate accounting of the voxel volume contributing to the spectrum being quantified, the transmitter and receiver radio-frequency field homogeneity, signal loss or distortion through partial saturation, T2 relaxation, *j* modulation, water suppression, and any Overhauser effect must be included (31,32). It is too easy to see how small systematic errors in each correction required to derive a concentration could accumulate to produce significant or large discrepancies over and above the claimed random errors in concentrations reported from different sites or studies (31).

Finally, one expects to see random scatter in measurements of metabolite concentrations or ratios that is commensurate with the particular S/N of the moieties in question. Because the percentage of uncertainty in the ratio of two metabolites is equal to the sum of the percentages of uncertainty in each, the choice of a metabolite with a low S/N for the quotient will provide measurements with larger scatter than normalization with a large peak. Low S/N may thus dictate that the error or standard deviation in measurements of, say, Pi and lactate in normal P-31 and H-1 MR spectra exceeds 50%. This does not necessarily mean that the results are useless or dismissable. If the error is truly random, a careful statistical analysis of many such measurements from different study groups (preferably blinded) can reveal a significant scientific or even clinically useful finding, depending on whether the difference between the means of each group is large enough and on the extent to which individual data overlap. Moreover, since each datum in the overlapping region corresponds to a patient whose spectroscopy findings are equivocal, the question of the everyday clinical utility of any spectroscopy test tends to pivot on the proportion of cases that exhibit such overlap.

DISTORTED SPECTRA

Distorted spectra and measurements derived therefrom pervade in vivo spectroscopy and are probably the major source of disparity among findings from different laboratories. The most common sources of distortion are T1 and T2 relaxation processes, the spatial localization technique employed, and water suppression techniques used in H-1 spectroscopy.

Different metabolites have different relaxation times (32,33), which results in differential partial saturation and/or signal decay when the repetition time is too short or echo time too long. If repetition time or echo time are not specified, the quantitative findings of a study may not be reproducible. The same is true of the flip angle of the excitation pulse that affects partial saturation.

If a surface coil is used to excite an MR signal with a nonuniform flip angle over the field of view, authors should explain their standardized protocol for adjusting the flip angle so that it is substantially the same in the region of interest for each of the examinations that comprise a clinical study. Alternatively, they should develop and validate a protocol for correcting the measurements for partial saturation, which is preferable. With such a diversity in repetition-time values and dubious flip-angle adjustments in P-31 studies especially, why should not investigators be responsible for somehow correcting their own metabolite values for partial saturation so that others might use them for comparison?

Examples of distortions caused by the localization technique used to acquire a spectrum are the baseline "wiggles" or "holes," sometimes resembling the wings of sinc functions, that surround resonances that are localized by means of methods employing imaging gradient magnetic fields like DRESS and phase-encoded gradient spectroscopy. These result from the small acquisition delay during which the gradients must be applied (34). Missing or attenuated peaks such as that of β-ATP in rotating frame techniques (15) are apparently due to bandwidth limitations. It may be too much to ask that measurements be corrected for such distortions as well, but authors must provide enough of the relevant information about sequence timing that skilled readers might gauge their affect on the results, if not perform the experiment themselves.

The famous "1-$\bar{3}$-3-$\bar{1}$" water-suppression pulse often employed for H-1 spectroscopy possesses frequency response characteristics with the capacity for enormous chemical shift–dependent spectral distortion (35), as illustrated in Figure 2. Chemical-selective pulses (36) offer inherently less distortion, but the advantage is easily lost if their widths are too narrow. The point here is that the absolute ratio of, say, *N*-acetyl aspartate or lactate to the total creatine pool or to choline could depend on how the water-suppression pulse was adjusted. Differences between study groups could arise from changes in the adjustment parameters or protocol.

VOLUMINOUS VOXELS

Less obvious and therefore more sinister are the effects of contamination of spectra from signals whose sources lie outside a declared voxel, which also hinge on how one defines voxel size (37). If the voxel is defined only by the amplitude of a sensitivity profile and it represents a small fraction of the volume of a sample, it can easily be swamped by the integrated signal from the bulk of that sample even if the amplitude of the signal at any point outside the voxel is tiny (5,37). For example, motion (37,38) and partial saturation when repetition time is less than T1 in ISIS (39) are potentially overwhelming sources of contamination, yet who performs ISIS with a repetition time equal to or greater than T1? Those who do not must describe their solutions to these problems.

In rotating frame and phase-encoded gradient spectroscopy, there is voxel "bleed" of signal intensity from adjacent voxels that depends on how much the signal sources vary within a voxel (37), but at least all of the artifact is distributed over a large expanse of the spectroscopic imaging array (as it is in MR imaging) and cannot easily find its way into a single "localized" spectrum. The point is that a spectrum claimed to be derived from, say, a 27-cm^3 voxel in a brain tumor cannot really be said to be derived from that tumor if 80% of the signal intensity in the spectrum is contamination from the rest of the brain (39).

It is also important to recognize that because even the smallest spectroscopy voxel of perhaps 1 cm^3 is manifold larger than those of MR imaging, it will

be the exception, not the rule, that voxels are contaminated by contributions from normal tissue and/or heterogeneous pathologic conditions identifiable with MR imaging. Conclusions, therefore, should either be bullet-proofed or tempered by the potentially confounding effects of tissue heterogeneity. In cases in which the spectral characteristics of one or possibly more of the tissue components that can be identified in a voxel are well defined, it may be possible to apply a correction to each spectrum to account for and substantially eliminate that contamination source.

For example, blood contamination of P-31 heart spectra in voxels intersecting the ventricular chamber can alter the observed PCr-to-ATP ratio, since blood contains ATP but no PCr. Since blood exhibits a characteristic DPG doublet with DPG/ATP $\approx$ 30%, its affect on the myocardial PCr/ATP measurement may be substantially compensated for by subtracting 15% of the total integrated DPG signal from the ATP integral for that voxel (40).

In a similar way, since liver and kidney contain no PCr, liver and kidney P-31 spectra that are contaminated by muscle signals might be corrected by subtracting from each resonance an amount of muscle signal contribution that is derived from the product of the known muscle metabolite ratios to PCr, with the measured amount of PCr contaminating the spectrum. Care in ensuring that the muscle spectra are in fact normal or at least do not vary between study groups, and that metabolite ratios assumed for the corrections are valid for the saturation conditions present during spectral acquisition (which may necessitate use of a uniform excitation field or a repetition time much greater than T1[PCr]), would be essential in avoiding other systematic errors in cases in which the corrections are large.

LOCALIZATION WARS

If spectroscopy were another anatomic imaging modality, only volume, section-selective, or "inner-volume" techniques of multivoxel spectroscopic image localization would probably now be acceptable. Adoption of this posture would dismiss the bulk of clinical research findings in spectroscopy to date that were discovered by other means (5). It would also ignore the fact that some pathologic conditions are better suited to particular localization strategies.

It would make little sense, for example, to employ spectroscopic imaging for a clinical examination of a muscular metabolic disease that globally affected the leg, when a simple surface receiver coil will do the trick (1,5). It would be nice to have a full three-dimensional phase-encoded P-31 spectroscopic image of the heart during stress testing of patients with coronary disease (16), but this presently cannot quite be accomplished in the 5 minutes or so allotted for spectroscopy acquisition during exercise on the basis of patient tolerance. Thus, a one-dimensional or some other three-dimensional hybrid (41) surface coil sequence is employed. If MR imaging precisely delimits a tumor, what is wrong with tailoring a single spectroscopy voxel to fit inside it and omitting acquisition of spectra from the remaining regions not of interest?

Witnesses to spectroscopy conferences still report duels between champions of localization techniques in search of that which is unflawed or at least less flawed, or that which is universal, yet a gaggle of different strategies remains productive. In light of the outstanding opportunities for voxel contamination and distortion in most in vivo situations as noted above, we might do well to paraphrase a parable: "Let he whose voxel is without contamination cast the first stone" when evaluating alternative localization approaches. What matters of course, is that the artifacts are under control and not overwhelming and that the experiment and its findings are reproducible and quantitative. The localization technique should be chosen to suit the pathologic condition and the patient limitations.

N

Studies composed of small patient populations, or larger ones that are so heterogeneous that only one or two pathologic conditions of any single type are represented, are problematic because whether the conclusions are characteristic of the pathologic condition in question remains unknown. Authors of papers about one or two patients who conclude little more than that one or two patients can be studied with a particular technology have not added significantly to the literature unless the innovation is found elsewhere in the paper. Papers that introduce and demonstrate a substantially new technology or a particularly unusual patient history or pathologic condition are examples of those in which small patient numbers might initially be acceptable. Otherwise, study populations should be sufficiently large to reflect the usual range of biologic heterogeneity in each group that is carefully defined in the Subjects section of the paper, and large enough to establish whether any statistically significant differences exist among the groups. This usually means performing identical measurements with a comparably sized control population.

Magnetic Resonance in Medicine now recommends in its information to authors of submissions that measurements from at least four subjects for communications, and 10 subjects for full papers, be included to establish significant biologic or medical conclusions. While I am opposed to journals setting rigid limits on the reporting of research, which is supposed to be innovative but not rigid, and although there is occasionally something to be said in a well-documented case study, it would not hurt to adopt these numbers as guidelines for those clinical spectroscopy research papers that possess few other novel features and are submitted to appear in *Radiology*.

APPENDIX

Suggested Guidelines for Description of Experimental Parameters in Clinical MR Spectroscopy and Chemical Shift Spectroscopic Imaging Studies

The following experimental parameters represent the consensus of a 1990–1991 Spectroscopy Advisory Subcommittee of the Society of Magnetic Resonance Imaging with respect to the parameters that describe in vivo spectroscopy and chemical shift imaging examinations. The subcommittee members were Cecil Charles, PhD, Kamil Ugurbil, PhD, Paul Bottomley, PhD, Truman Brown, PhD, Robert Lenkinski, PhD, Ian Young, PhD, Wolfhard Semmler, MD, PhD, and Michael Weiner, MD (chairman).

I. Equipment Parameters
  A. Magnet
    1. Static field strength
       The field strength is expressed in tesla.
    2. Homogeneity adjustment
       The means of homogeneity adjustment (shimming), including any localization technique, should be specified, and the homogeneity obtained in the given experiment should be reported as frequency width half maximum in parts per million or hertz for the observed tissue.
  B. Radio-Frequency Field (B_1)
    1. Frequency of observed nucleus
    2. Radio-frequency coil characteristics
       a) physical/electrical/MR characteristics
       b) size and shape
       c) special designs (eg, coplanar arrays, other arrays, volume/surface coil combinations, conformal coils, and packaging where appropriate, as in insertable coils)
  C. Gradient Fields
     Describe the relevant gradient waveform(s) and strength.
II. Acquisition Technique
  A. Localization Technique
    1. Reference or description of localization technique
    2. Image reference (if appropriate)
    3. Solvent suppression for hydrogen studies: The technique and the water-suppression factor achieved in the experiment should be stated.

B. Sampling Technique
If nonlinear techniques are used, they should be described or referenced and the sampling criterion noted.

C. Artifact Reduction Methods
Methods used to minimize artifacts, such as cardiac gating and gradient reordering, should be described or referenced.

D. Scan Parameters (underlined parameters strongly recommended)
1. *Repetition time*
2. *Echo time* (if applicable)
3. *Inversion time* (if applicable)
4. *Mixing period* (in a two-dimensional spectroscopy experiment, if applicable)
5. *Flip angle: The procedure for setting flip angle should be described.*
6. *Width of nonselective pulses*
7. Selective pulses
   a) *pulse width*
   b) bandwidth
   c) additional descriptors: *shape* (mathematical description/graphic description), *window functions* (eg, Hanning), number of data points in waveform
8. Spectral width
9. Field of view
10. Number of acquired points
11. Number of reconstructed points (see below)
12. *Whether free induction decay, half echo, or whole echo is acquired*
13. Receiver dead time delay (optional)
14. *Additional delays in free-induction-decay acquisitions due to hardware limitations*
15. Number of excitations/encodings/etc
16. *Total imaging time* (including any dummy acquisitions)
17. *Total examination time* (including coil placement/shimming/data acquisition)

III. Data Reconstruction, Display, and Analysis

A. Reconstruction
1. Fourier transform or other algorithm
2. Zero padding or other interpolation
3. Filter functions used in the spatial and/or chemical shift domains
4. Phase correction
   a) orders of correction used (zero, first, higher)
   b) whether manual or automatic (algorithm should be noted)
5. Baseline correction: For all baseline corrections involving adjustment of selected points to zero, the means of selecting these points should be explained.
   a) manual (eg, linear tilts)
   b) automatic: Algorithm (eg, cubic spline) should be noted.

B. Display
1. Spectra
   a) Chemical shift scale should be present with a chemical shift reference identified.
   b) The means of peak identification should be indicated (eg, from literature reference or titration with standard).
   c) Stack plots/contour plots: The additional dimension (time, space, etc) should be clearly labeled with appropriate scalars, and for contour plots, the contour scale should be displayed with the data.
2. Metabolic images (chemical shift imaging/spectroscopic imaging)
   a) Images of specific chemical shift resonances should have spatial labels (right/left, anterior/posterior, superior/inferior) for coordination with a high-resolution image; label(s) of the species (eg, lactate, Pi); any interpolation technique used for the image; the image depth (frequency width) in the chemical shift dimension; the spectral mode displayed (eg, real, imaginary, or magnitude component) and an intensity scale.
   b) Images that are generated from curve-fitted data rather than raw data should be so labeled.

C. Analysis
1. Curve fitting (time or frequency)
   a) Fitting algorithm should be described or referenced.
   b) An error analysis should be included.
2. A signal-to-noise analysis of spectra should be included. ■

ADDENDUM

The question of the correct value for the PCr/ATP ratio in normal human heart tissue raised in the Table was addressed at a discussion forum held at the 10th Annual Society of Magnetic Resonance in Medicine Meeting in San Francisco, August 10–16, 1991. Panelists J. S. Ingwall, PhD, P. Styles, PhD, S. Schaefer, MD, P. R. Luyten, PhD, P. M. L. Robitaille, PhD, and I evaluated existing evidence and reached a consensus of opinion that the true value likely falls in the range of 1.6–2.0. As some issues resolve, however, new ones breed at the frontier, such as the consistency with which lactate detected with H-1 spectroscopy can be induced in the human visual cortex by means of visual stimulation (42,43): Analogous experiments with positron emission tomography show increased uptake of fluorodeoxyglucose with visual stimulation.

References

1. Ross BD, Radda GK, Gadian DG, Rocker G, Esiri M, Falconer-Smith J. Examination of a case of suspected McArdle's syndrome by ^{31}P nuclear magnetic resonance. N Engl J Med 1981; 304:1338–1342.
2. Cooper LS, Chalmers TC, McCally M, Berrier J, Sacks HS. The poor quality of early evaluations of magnetic resonance imaging. JAMA 1988; 259:3277–3280.
3. Margulis AR. About articles of judgment. Radiology 1988; 169:576–577.
4. Shuman WP, Hillman BJ, Friedman PJ, Rosenquist CJ, Stark DD. Early evaluations of MR imaging: worthless or worthwhile? AJR 1988; 15:857–864.
5. Bottomley PA. Human in vivo NMR spectroscopy in diagnostic medicine: clinical tool or research probe? Radiology 1989; 170:1–15.
6. Fossel ET, Carr JM, McDonagh J. Detection of malignant tumors: water suppressed proton nuclear magnetic resonance spectroscopy of plasma. N Engl J Med 1986; 315:1369–1376.
7. Engan T, Krane J, Klepp O, Kvinnsland S. Proton magnetic resonance spectroscopy of plasma from healthy subjects and patients with cancer. N Engl J Med 1990; 322: 949–953.
8. Okunieff P, Zietman A, Kahn J, et al. Lack of efficacy of water-suppressed proton nuclear magnetic resonance spectroscopy of plasma for the detection of malignant tumors. N Engl J Med 1990; 322:953–958.
9. Eugene M, Le Moyec L, de Certaines J, et al. Lipoproteins in heart transplantation: proton magnetic resonance spectroscopy of plasma. Magn Reson Med 1991; 18:93–101.
10. Pont H, Vion-Dury J, Kriat M, et al. NMR spectroscopy of plasma during acute rejection of transplanted hearts. Lancet 1991; 337:792–793.
11. Oberhaensli RD, Galloway GJ, Hilton-Jones D, et al. The study of human organs by phosphorus-31 topical magnetic resonance spectroscopy. Br J Radiol 1987; 60:367–373.
12. Hubesch B, Sappey-Marinier D, Roth K, Meyerhoff DJ, Matson GB, Weiner MW. P-31 MR spectroscopy of normal human brain and brain tumors. Radiology 1990; 174:401–409.
13. Luyten PR, Groen JP, Vermeulen JWAH, den Hollander JA. Experimental approaches to image localized human ^{31}P NMR spectroscopy. Magn Reson Med 1989; 11:1–21.
14. Bottomley PA, Herfkens RJ, Smith LS, Bashore TM. Altered phosphate metabolism in myocardial infarction: P-31 MR spectroscopy. Radiology 1987; 165:703–707.
15. Blackledge MJ, Rajagopalan B, Oberhaensli RD, Bolas NM, Styles P, Radda GK. Quantitative studies of human cardiac metabolism by P-31 rotating frame NMR. Proc Natl Acad Sci USA 1987; 84:4283–4287.
16. Weiss RG, Bottomley PA, Hardy CJ, Gerstenblith G. Regional myocardial metabolism of high-energy phosphates during isometric exercise in patients with coronary artery disease. N Engl J Med 1990; 323: 1593–1600.
17. Bottomley PA, Weiss RG, Hardy CJ, Baumgartner WA. Myocardial high-energy phosphate metabolism and allograft rejection in patients with heart transplants. Radiology 1991; 181:67–75.

18. Bottomley PA, Hardy CJ, Roemer PB. Phosphate metabolite imaging and concentration measurements in human heart by nuclear magnetic resonance. Magn Reson Med 1990; 14:425–434.
19. Schaefer S, Gober J, Valenza M, et al. Nuclear magnetic resonance imaging-guided phosphorus-31 spectroscopy of the human heart. J Am Coll Cardiol 1988; 12:1449–1455.
20. Luyten PR, Bruntink G, Sloff FM, et al. Broadband proton decoupling in human ^{31}P NMR spectroscopy. NMR Biomed 1989; 1:177–183.
21. Schaefer S, Gober JR, Schwartz GG, Twieg DB, Weiner MW, Massie B. In vivo phosphorus-31 spectroscopic imaging in patients with global myocardial disease. Am J Cardiol 1990; 65:1154–1161.
22. Bottomley PA, Hart HR, Edelstein WA, et al. Anatomy and metabolism of the normal human brain studied by magnetic resonance at 1.5 Tesla. Radiology 1984; 150: 441–446.
23. Bottomley PA, Hardy CJ, Cousins JP, Armstrong M, Wagle WA. AIDS dementia complex: brain high-energy phosphate metabolite deficits. Radiology 1990; 176:407–411.
24. Menon DK, Baudouin CJ, Tomlinson D, Hoyle C. Proton MR spectroscopy and imaging of the brain in AIDS: evidence of neuronal loss in regions that appear normal with imaging. J Comput Assist Tomogr 1990; 14:882–885.
25. Luyten PR, Marien AJH, Heindel W, et al. Metabolic imaging of patients with intracranial tumors: H-1 MR spectroscopic imaging and PET. Radiology 1990; 176:791–799.
26. Madden A, Leach MO, Sharp JC, Collins DJ, Easton D. A quantitative analysis of the accuracy of in vivo pH measurements with ^{31}P NMR spectroscopy: assessment of pH measurement methodology. NMR Biomed 1991; 4:1–11.
27. Dumoulin C. Automatic peak assignments for in vivo ^{31}P NMR spectra. Comput Enhanced Spectroscopy 1984; 2:61–67.
28. Nelson SJ, Brown TR. The accuracy of quantification from 1D NMR spectra using the PIQABLE algorithm. J Magn Reson 1989; 84:95–109.
29. Spielman D, Webb P, Macovski A. A statistical framework for in vivo spectroscopic imaging. J Magn Reson 1988; 79:66–77.
30. van der Veen JWC, de Beer R, Luyten PR, van Ormondt D. Accurate quantification of in vivo ^{31}P NMR signals using the variable projection method and prior knowledge. Magn Reson Med 1988; 6:92–98.
31. Bottomley PA, Hardy CJ. Rapid, reliable in vivo assays of human phosphate metabolites by nuclear magnetic resonance. Clin Chem 1989; 35:392–395.
32. Roth K, Hubesch B, Meyerhoff DJ, et al. Noninvasive quantitation of phosphorus metabolites in human tissue by NMR spectroscopy. J Magn Reson 1989; 81:299–311.
33. Merboldt KD, Chien D, Hanicke W, Gyngell ML, Bruhn H, Frahm J. Localized ^{31}P NMR spectroscopy of the adult human brain in vivo using stimulated-echo (STEAM) sequences. J Magn Reson 1990; 89:343–361.
34. Allman T, Holland GA, Lenkinski RE, Charles HC. A simple method for processing NMR spectra in which acquisition is delayed: applications to in vivo localized ^{31}P NMR spectra acquired using the DRESS technique. Magn Reson Med 1988; 7:88–94.
35. Morris GA, Smith KI, Waterton JC. Pulse sequences for solvent suppression with minimal spectral distortion. J Magn Reson 1986; 68:526–532.
36. Bottomley PA, Edelstein WA, Foster TH, Adams WA. In vivo solvent-suppressed localized hydrogen nuclear magnetic resonance spectroscopy: a window to metabolism? Proc Natl Acad Sci USA 1985; 82: 2148–2152.
37. Bottomley PA, Hardy CJ, Roemer PB, Weiss RG. Problems and expediencies in human ^{31}P spectroscopy: the definition of localized volumes, dealing with saturation and the technique-dependence of quantification. NMR Biomed 1989; 2:284–289.
38. Bottomley PA, Hardy CJ. PROGRESS in efficient three-dimensional spatially localized in vivo ^{31}P NMR spectroscopy using multidimensional spatially selective (ρ) pulses. J Magn Reson 1987; 74:550–556.
39. Lawry TJ, Karczmar GS, Weiner MW, Matson GB. Computer simulation of MRS localization techniques: an analysis of ISIS. Magn Reson Med 1989; 9:299–314.
40. Hardy CJ, Weiss RG, Bottomley PA, Gerstenblith G. Altered myocardial high-energy phosphate metabolites in patients with dilated cardiomyopathy. Am Heart J 1991; 122:795–801.
41. Hardy CJ, Bottomley PA. ^{31}P spectroscopic localization using pinwheel NMR excitation pulses. Magn Reson Med 1991; 17:315–327.
42. Prichard J, Rothman D, Novotny E, et al. Lactate rise detected by ^{1}H NMR in human visual cortex during physiologic stimulation. Proc Natl Acad Sci USA 1991; 88:5829–5831.
43. Merboldt KD, Bruhn H, Gyngell ML, Hanicke W, Michaelis T, Frahm J. Variability of lactate in normal human brain in vivo: localized proton MRS during rest and photic stimulation (abstr). In: Book of abstracts: Society of Magnetic Resonance in Medicine 1991. Berkeley, Calif: Society of Magnetic Resonance in Medicine, 1991; 392.