AJNR

Is Quality Medical Care Dead or Just Buried Alive?

James N. Dreisbach

AJNR Am J Neuroradiol 2000, 21 (8) 1375-1376
http://www.ajnr.org/content/21/8/1375

MEG versus BOLD MR Imaging: Functional Imaging, the Next Generation?

Blood oxygen level–dependent functional MR (BOLD fMR) imaging, long accepted as a powerful research technique in the cognitive sciences and neurosciences, recently has gained acceptance as a clinical tool. Clinical applications of BOLD imaging have focused primarily on the preoperative localization of the motor, sensory, and language centers of the brain in anticipation of tumor or vascular malformation resection, and in the functional evaluation of focal cortical dysplasias in epilepsy. Neuronal firing is not directly measured by fMR imaging. Rather, in BOLD imaging, as its name implies, neuronal activation is inferred from small, local MR signal changes (on the order of 3% at 1.5 T field strength) proportional to hemodynamically induced alterations in net deoxyhemoglobin concentration caused by task-related increases in neuronal metabolism. Thus, although capable of spatial resolution on the order of millimeters, BOLD fMR imaging is limited in its temporal resolution to the time interval required for a change in neuronal activity to produce a measurable hemodynamic response, typically approximately 2 to 5 seconds.

Magnetoencephalography (MEG), on the other hand, a somewhat newer technique, not only directly measures the magnetic field changes associated with neuronal firing, but is capable of temporal resolution on the order of milliseconds. This high temporal resolution is likely to prove especially valuable in tracking transient, complex, coordinated neuronal activation patterns involved in higher-order cognitive functions (such as visual memory formation), which are known to occur across large segments of the brain (1). In MEG, an array of hypersensitive, superconducting magnetic detectors translate minute, rapidly changing magnetic fields (on the order of one billionth the strength of the earth's magnetic field) into detectable alterations in electric current. This is accomplished, however, at the price of decreased spatial resolution compared to that of fMR imaging. The spatial resolution of MEG, for unspecified magnetic source distributions, is typically limited to several centimeters.

In this issue of the *AJNR*, Roberts et al (page 1377) report on the difference between the ability of BOLD fMR imaging and MEG to quantify evoked responses. Specifically, they studied five subjects, all of whom underwent a similar sensory paradigm, using both techniques. The ''task'' involved successive stimulation of one, two, three, and four digits of the left hand. For fMR imaging, activation was quantified in two ways: first, as the extent of cortical activation, and second, as the amount of activation (defined as the product of the number of activated pixels and the mean signal change per pixel). For MEG, activation was also quantified in two ways: first, as the magnitude of the evoked magnetic field peak, and second, as the strength of the modeled current source, Q. Using fMR imaging, a trend toward an increased number of activated pixels for an increased number of stimulated digits was not found to be statistically significant, and a very high intrasubject and intersubject variability in pixel activation was noted. Using MEG, however, the evoked field magnitude was found to vary linearly with the number of digits activated in a statistically significant way, and intrasubject and intersubject variability of activation was noted to be much less than it had been for fMR imaging. The authors concluded that, for the particular somatosensory task they studied, robust quantification of evoked responses is possible with MEG, but robust quantification of increasing cortical area of activation is not possible with fMR imaging.

In attempting to assess the clinical impact of these results, two questions present themselves. The first is a question of methodology; were alternative experimental designs possible? Did this study compare BOLD and MEG so as to optimize the possibility of quantitation for each technique, and are the results reproducible? The second is a question of relevance; are the conclusions of this study important? Does quantification of cortical activation examinations matter clinically, and what are the potential roles for MEG and fMR imaging in patient care?

With regard to the first question, alternative experimental designs might have cast a more favorable light on the potential of fMR imaging to document reproducible, quantifiable, evoked responses. The conclusion reached by Roberts et al holds strictly only for the particular paradigm that they studied. It is possible that by varying the frequency, duration, intensity, or even the type of stimulus presentation, rather than varying only the number of digits stimulated, the resulting fMR imaging signal changes might have been more robust to quantification (or the MEG changes less). The way in which quantification is defined, which was different for the two techniques studied, could also affect the results. Because robust hemodynamic alterations are detectable after neuronal stimuli lasting only a few 10s of milliseconds, a newer method of BOLD imaging, ''event related'' fMR imaging, has been designed to measure activation in response to single sensory or cognitive events (2). Had this study used an ''event-related'' rather than a ''block'' design for its stimulus presentation, or had the ''power'' of the fMR imaging response (the area under the signal intensity versus time curve for an activated pixel) been measured, rather than the extent and amount of activa-

tion (as defined by Roberts et al), the results might have been different. fMR imaging data collection using a magnetic field strength greater than 1.5 T (3 T or higher) also could have improved the signal-to-noise ratio of the BOLD effect.

With regard to the second question, accurate, reproducible quantitation of functional activation could be valuable not only to monitor serially the response of individual patients to treatment in situations such as stroke rehabilitation, but also as an objective surrogate marker of disease progression when comparing clinical trial outcomes among different subjects. For such applications, a low intrasubject and intersubject variability is required. For preoperative planning, although localization of function is crucial, quantitation of evoked responses might help determine the significance of equivocal foci of activation, which may be present in regions of peritumoral edema or mass effect (3). In the future, MEG also may prove useful in quantifying disease states such as epilepsy, memory disorders (dementias), or language disorders, for which structural derangement might not be apparent on conventional MR imaging.

An important drawback in the use of MEG to evaluate diseases such as epilepsy, however, (despite its great sensitivity for detecting abnormal ''spikes'' of neuronal activation) is that anatomic localization with MEG is highly model dependent. When, as in the study by Roberts et al, the approximate brain region of an evoked response is known with high certainty (the postcentral gyrus, in the case of somatosensory stimulation), the models used to infer the strength and location of the current source, Q, can be considered reliable. When the approximate location of a current source is less certain, or the sources are multiple and widely distributed over the cortical surface, anatomic localization with MEG can be considerably less reliable.

Given the preceding discussion, it is clear that important differences exist between fMR imaging and MEG in their sensitivity for detection of evoked responses, how they measure such responses, and what those responses mean. MEG and fMR imaging techniques have largely complementary strengths and weaknesses, not only with respect to their ability to quantify functional activation, but also in their spatial and temporal resolutions. Rather than emphasize the limitations of each of these techniques, however, it makes sense to attempt to exploit their strengths in order to optimize their clinical and research utility. To this end, techniques recently have been developed that combine the spatial resolution of fMR imaging with the temporal resolution of MEG to create ''anatomically constrained'' functional activation maps (4). These maps are calculated by using the high resolution fMR imaging and structural MR imaging datasets, obtained during activation studies, to model more precisely the highly temporally resolved MEG current sources. The results are superimposed onto an ''inflated'' cortical surface representation, and can be displayed as ''movies'' showing spreading waves of cortical activation.

Using such ''anatomically constrained'' maps, it is possible to obtain highly spatially and temporally resolved functional information that is unavailable from data produced by either BOLD fMR imaging or MEG alone. This technique has been used successfully to study parallel cortical activation during semantic processing of visually presented words in diverse brain regions associated with perception, semantic processing, and response choice (4). By building on the complementary strengths of fMR imaging and MEG, precise monitoring of the spatiotemporal orchestration of complex, high-order, perceptual and cognitive neuronal activations has the potential to become a clinically useful technique for the evaluation of both normal and diseased states.

MICHAEL H. LEV, M.D.
P. ELLEN GRANT, M.D.

Massachusetts General Hospital
Boston, MA

References

1. Glanz J. **Magnetic brain imaging traces a stairway to memory.** *Science* 1998;280:37
2. Rosen BR, Buckner RL, Dale AM. **Event-related functional MRI: past, present, and future.** *Proc Natl Acad Sci* 1998;95: 773–780
3. Holodny AI, Schulder M, Liu WC, Maldjian A, Kalnin AJ. **Decreased BOLD functional MR activation of the motor and sensory cortices adjacent to a glioblastome multiforme: implications for image-guided neurosurgery.** *AJNR Am J Neuroradiol* 1999;20:609–612
4. Dale AM, Liu AK, Fischl B, et al.**Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity.** *Neuron* 2000;26:55–67

Task-correlated Head Movement in fMR Imaging: False Activations Can Contaminate Results Despite Motion Correction

In recent years, functional magnetic resonance (fMR) imaging has greatly expanded our capacity to investigate the neuronal substrates of human cognitive processes. This methodology, which relies on blood oxygenation level–dependent (BOLD) contrast, has proven to be a valuable tool for addressing not only questions regarding the basic nature of human cognitive function, but also questions concerning how aging and disease can alter this function.

As in all scientific endeavors, neuroimaging is susceptible to errors. A great deal of attention has been devoted to avoiding statistically false-positive results. That is, given that independent statistical tests are conducted on thousands of voxels, and that each test has a small probability of falsely concluding that there is a significant activation, the sheer number of tests results in a large number of brain regions that spuriously appear to be activated.

False-positive results can also occur when head motion is correlated with the task design, and it is this type of artifact that Field and colleagues address in their article in this issue of the *AJNR* (page 1388). As an example, if a study investigates brain regions underlying movement of the fingers, and employs an experimental protocol in which periods of finger movement are alternated with periods of rest, it is possible that the subject's finger movements will translate subtle motion to the head during the finger movement blocks. Such motion manifests as translation along or rotation about the *x*, *y*, or *z* axis, and can produce regional differences in signal magnitude between two contrasting conditions (movement vs rest in this example) that reach statistical significance. In contrast to statistically false-positive results (ie, type I errors), it is difficult to assess the probability of motion-induced false-positive results. Most researchers strive to minimize this probability either ''on-line'' by reducing the opportunity for head motion in the first place by using bite-bars or other head restraint devices, ''off-line'' by using motion correction postprocessing algorithms to realign all the brain volumes to a reference volume, or by a combination of these methods.

Although previous investigators (1) have demonstrated that relatively large movements (approximately 3-mm translation) can result in spurious activations that are reduced by motion correction algorithms, the investigation by Field et al is unique in three ways. First, a phantom approximating the size and shape of the human brain was constructed, along with an apparatus for introducing controlled in-plane translations and rotations. Thus, simulations of fMR imaging experiments with alternating blocks of two different trial types could be performed and, in contrast to studies using human volunteers, task-correlated motion could be guaranteed to be present while task-correlated neuronal activation was guaranteed to be absent. Second, the effects of subtle movements (< 1 mm) with varying degrees of task-correlated motion were investigated to simulate realistic experimental conditions. Finally, false-positive results due to motion were assessed after employing sophisticated postprocessing algorithms, including motion correction, removal of low-frequency components, and corrections for multiple comparisons using spatial extent. These analytical methods are commonly employed in fMR imaging investigations.

Field et al observed that, despite the subtlety of movement and the inclusion of accepted postprocessing procedures, false activation appeared when movement correlated with the task at $r > 0.52$, and appeared on every experiment with $r > 0.67$. The authors argue that ''the degree of correlation between stimulus and motion may be more important than the magnitude of motion in creating these artifacts.'' Although the investigation of Field et al has methodological limitations (eg, the phantom has a different structural and chemical composition than the human brain, which could result in relatively greater sensitivity to motion-related artifacts), their results should nevertheless raise concerns within the neuroimaging community about the degree to which motion contributes to fMR imaging activation maps.

Although Field et al have increased awareness that the potential for motion-related false-positive results may be present even when motion has been ''prevented'' or ''corrected,'' their results raise a number of questions:

- How should investigators modify their procedures to reduce the probability of motion-related false-positive results? At a minimum, it seems reasonable to suggest that investigators monitor the magnitude of correlation between the task and motion for each subject. The most conservative approach would be to discard subjects with unacceptably high correlations, but other corrective measures may be possible and deserve further attention.
- What is the effect of motion on false-negative results? That is, how often does subtle motion eliminate or reduce genuine neuronally derived activation?
- Is through-plane movement more or less likely to produce artifacts than in-plane movement? Field et al investigated only in-plane movement.
- Are event-related fMR imaging investigations less susceptible to motion-related artifacts than block designs? Field et al simulated a block design with six alternating ''on'' and ''off'' epochs of 30 seconds each.

Further investigation of these matters will likely improve the quality of functional neuroimaging data, and will increase confidence that results reflect genuine activation rather than motion.

JOHN E. DESMOND, PH.D.
SCOTT W. ATLAS, M.D.
Stanford University, Stanford, CA

References

1. Hajnal JV, Myers R, Oatridge A, Schwieso JE, Young IR, Bydder GM. **Artifacts due to stimulus correlated motion in functional imaging of the brain.** *Magn Reson Med* 1994;31:283–291

Aneurysm Rupture during GDC Treatment: Optimizing the Rescue Strategy

In the article *Use of a Second Microcatheter in the Management of a Perforation during Endovascular Treatment of a Cerebral Aneurysm* in this issue of the *AJNR* (page 1537), Willinsky et al present a novel and rational modification of the accepted strategy for managing a complication all too familiar to most interventional neuroradiologists. The authors left the initial microcatheter traversing the aneurysm dome while a second microcatheter was used to access the aneurysm and coil the lumen. The patient tolerated the event without any adverse sequelae.

Since its development in 1990, Guglielmi detachable coil (GDC) embolization of intracranial aneurysms has evolved from an experimental procedure to a well-accepted and widely performed method of protecting patients from subarachnoid hemorrhage. More than 30 000 intracranial aneurysms have been treated worldwide, and the variables that affect the procedure's safety and efficacy are well documented. In considering the procedure's safety, it is worth noting that the two most common complications reported are thromboembolic and hemorrhagic, with the former significantly more common than the latter.

Intraprocedural aneurysm hemorrhage can be spontaneous or caused by overdrainage of a ventricular drain catheter, but more commonly might be caused by contrast overinjection or wire, catheter, or coil perforation. Of these, forward migration of the microcatheter is the most common cause of aneurysm rupture during GDC embolization. This emphasizes the importance of safe microcatheterization of aneurysms—reducing the risk for microcatheter jump by vigilant monitoring during microcatheter positioning, so that the forward progress of the catheter tip is commensurate with the forward progress of its shaft proximally. As a rule, the longer, the smaller in diameter, and the more tortuous the segment to be traversed by the microcatheter (from guide catheter tip to aneurysm) is, the more sites there are for friction between the microcatheter and vessel wall to accumulate (and suddenly release). One technique helpful in making microcatheter advancement more controlled is the triaxial technique. This technique, using a Tracker 38 as a middle catheter, effectively minimizes the potential build-up of friction, reducing the risk of microcatheter jump. In addition to these considerations, certain microcatheters are inherently more ''jumpy'' by virtue of their design characteristics. Braided catheters (including the microcatheter used in this case) are less likely to jump than nonbraided catheters.

Although microcatheter perforation into the subarachnoid space warrants an aggressive response, it must be kept in mind that not all cases of microcatheter tip migration beyond the confines of the aneurysm lumen represent perforation—catheters can migrate into the thrombosed portion of an aneurysm. In addition, not all aneurysm perforations are into the subarachnoid space. In the article by Willinsky et al, it appears at least possible that the perforation of the paraophthalmic aneurysm was into the cavernous sinus. The angiographic views provided suggest that this aneurysm had an intradural neck, but a dome extending into the cavernous sinus. These aneurysms can present with subarachnoid hemorrhage from the neck. The image of the microcatheter protruding through the dome shows a downward course. The patient suffered a very brief episode of mild hypertension, but did not display the classic Cushing hemodynamic response. Postprocedure CT images showed no new subarachnoid hemorrhage. The point is to confirm a subarachnoid location of the perforating catheter tip with an injection of a tiny amount of contrast, which would also confirm that reversal of heparin (with risks of clot formation, as occurred in this case) is truly necessary.

When aneurysm perforation into the subarachnoid space has occurred, quick response can salvage an ominous situation. Reversing the systemic heparin, achieving dense GDC packing of the aneurysm, and (if clinical signs of increased intracranial pressure or CT findings warrant) placing a ventricular drain constitute a strategy that can result in excellent neurologic recovery from an angiographically frightful hemorrhage. In describing the technique of leaving one microcatheter in place across the perforation while using a second catheter to coil the aneurysm, Willinsky et al have contributed a concept to our field that can improve the outcome of our patients.

An alternative strategy on the horizon is the use of liquid embolic agents. One such agent (Onyx, Micro Therapeutics, Inc) is currently being tested in clinical trials for aneurysm embolization. Liquid agents may be superior in the management of acute perforations and ruptures because, unlike GDC coils, these agents appear to seal the aneurysm lumen immediately.

Finally, it may be worthwhile to have a dialog within our specialty regarding the strategy of placing a coil across the perforation site, leaving it to transfix the wall of the aneurysm dome. As noted by Willinsky et al, this technique has been advocated in the management of intracranial vascular perforations. One must question, however, whether such placement of a coil is likely to seal the defect in a case such as the one they report (the defect represented a tear in a thin, abnormal aneurysm wall and was at least the diameter of the catheter—significantly larger than the diameter of the coil). One can also question whether the coil may impede healing of the perforation. Certainly there is a risk that the coil will exert tension on the edge of the tear because the coil itself is affected by blood flow

within the aneurysm, CSF flow outside the aneurysm, and gravity. Unfortunately, there is no experimental or clinical data upon which one can base an answer to this question with respect to intracranial aneurysms. At present, the question of whether a GDC coil left transfixing an aneurysm wall defect improves or hinders hemostasis and healing is left to our intuition. It is an issue warranting objective investigation.

TIM W. MALISCH, M.D.
Northwestern University Medical School
Chicago, IL

Cementing the Evidence: Time for a Randomized Trial of Vertebroplasty

Percutaneous vertebroplasty is a technique for treating low back pain that appears to be rapidly disseminating throughout the United States, and now O'Brien et al (page 1555) add their series of six patients to the literature in this issue of the *AJNR*. Yet, there are still no randomized, controlled trials that compare the long-term outcomes of percutaneous vertebroplasty to a control therapy.

Is it too late for a randomized trial? At last year's meeting of the American Society of Neuroradiology, it was suggested that a randomized trial for vertebroplasty would be unethical because patients would be denied the obvious benefit derived from the technique. In 1997, the World Medical Association issued the Declaration of Helsinki, which contained recommendations for physicians using human subjects in medical research (1). This declaration states, ''In any medical study, every patient —including those of a control group, if any, should be assured of the best proven diagnostic and therapeutic method. This does not exclude the use of inert placebo in studies where no proven diagnostic or therapeutic method exists.'' But this begs the question, what constitutes necessary and sufficient evidence to prove the efficacy of a therapy? After all, it is equally wrong to advocate the use of a new therapy that has not been shown to be more beneficial than standard treatment as it is to withhold an unproven therapy. Relying on anecdotal case reports and case series may lead to erroneous and harmful conclusions.

The power of modern Western medicine is derived in great part from its close alliance to the world of science (2) by using the scientific method to distinguish what is useful from what is not. Sir William Osler said, ''The philosophies of one age have become the absurdities of the next. . . '' (3). The history of medicine is littered with examples of treatments that went unquestioned, yet now provoke amusement or even horror. In the 16th century, surgeons treated gunshot wounds by pouring burning oil over them (4), until Ambroise Pare ran out of oil during an assault on Turin in 1537. He improvised an emulsion of eggs, rosewater, and turpentine and discovered that its use caused less swelling, the patients suffered less, and fewer patients died than when he had treated with the boiling oil. In the 19th century, the medical profession gained tremendous authority by adopting the scientific method to determine the value of medical practices. Empirical evidence showed that commonly accepted practices such as bloodletting had no therapeutic value (2). Simultaneously, emerging disciplines such as bacteriology and epidemiology began to benefit the health of the public in an indisputable and quite visible manner. The marriage between medicine and science was proving to be a great success.

As history has revealed, simple conviction that a treatment works can be horribly misleading. Such mistakes are not relegated to prior centuries. Several reviews of modern medical practice illustrate treatments that were accepted as standard and beneficial, but were found to be useless or even harmful when evaluated by a randomized trial. Although the use of empirical evidence to justify medical practice is a powerful principle, the quality of evidence for making medical decisions has been and continues to be highly variable. Using poorly acquired or incomplete evidence can result in disastrous decisions, and one does not have to look far back in history to find troubling incidents.

Perhaps the most notorious recent example is that of the antiarrhythmics encainide, flecainide, and moricizne. Their story is chronicled in the book *Deadly Medicine* by Thomas Moore (5). In the early 1980s these newly introduced antiarrhythmics were found to be highly successful at suppressing arrhythmias. On this basis, these drugs were widely promoted and commonly prescribed. Not until a randomized, controlled trial that looked at the ultimate outcomes of patients was performed was it realized that, although these drugs suppressed arrhythmias, they actually increased mortality. The Cardiac Arrhythmia Suppression Trial revealed that postmyocardial infarction patients with mild arrhythmias who were put on these drugs had an excess mortality of 56/1000. By the time the results of this trial were published, at least 100 000 such patients had been taking these drugs (5), meaning that 5600 people had died per year because of these agents.

Although there are numerous factors responsible for this medical mishap, the most important lesson that applies to vertebroplasty is that it is dangerous to rely on surrogate outcomes when assessing the benefit of a medical intervention. Proponents of these antiarrhythmics based their favorable impressions on the ability of the drugs to suppress arrhythmias. That arrhythmia suppression would de-

crease mortality was a reasonable biological hypothesis, but it proved erroneous in the end. Similarly, it seems reasonable that preventing further vertebral body collapse or even possibly restoring height might reduce the pain associated with osteoporotic compression fractures, but this remains to be proven. Fleming pointed out that surrogate outcomes frequently fail to predict the clinical outcome of interest (6). More importantly, the intervention ''might also affect the clinical outcome by unintended, unanticipated, and unrecognized mechanisms of action that operate independently of the disease process.'' Fleming concluded that, except for rare circumstances, surrogate outcomes should be avoided for definitive phase three trials.

While short-term pain relief augurs well for long-term benefits, no well-controlled study has shown even this short-term benefit. There is the distinct possibility that these short-term benefits will not last, and that in the long run, patients who undergo vertebroplasty might do no better or even worse than a control cohort.

Are case series adequate evidence to form a conclusive opinion? Although they are valuable for providing preliminary evidence, case series are rarely sufficient for making major medical policy decisions. There are several reasons why case series may be misleading when studying low back pain treatments (7). First, the natural history of acute low back pain in general, and the pain associated with osteoporotic compression fractures specifically, is to improve, usually regardless of the type of therapy. Part of this improvement reflects ''regression to the mean.'' This is a statistical concept that emphasizes that extreme values at one measurement of a variable tend to regress back toward a mean value when measured again. Patients with back pain tend to seek care when their pain is extreme. Regression to the mean implies that when such patients are seen on a follow-up visit, their pain will have improved (regressed to some average level), regardless of interventions. Second, because case series do not have a control group, the placebo effect may play a role in improvement. This effect applies not only to the technique of vertebroplasty, but also to the powerful influence of the enthusiasm and conviction of the physician performing vertebroplasty (8).

The ethical basis for conducting randomized trials relies on the uncertainty as to whether the intervention will be beneficial or harmful. If there is no uncertainty, then there is no need for a trial. If uncertainty does exist, however, then not only is it ethical to perform a trial, but it is necessary to choose the methodology most likely to eliminate the uncertainty. Proponents of a new technology that has been disseminated before it has been rigorously evaluated commonly argue that scientific evaluation would be unethical. In the example of antiarrhythmics cited above, many proponents thought that controlled trials would be unethical because these drugs were effective at suppressing arrhythmias (5). Dixon pointed out that this kind of specious argument is predictable and standard (9). He argued that social forces are more important than scientific forces in determining clinical policies, and that characteristic errors occur in the formation of these policies. One of these errors is the defense of unproven, prematurely disseminated technologies with the argument that it is unethical to stop and rigorously evaluate them.

Nonetheless, ethics insists that we do stop. Vertebroplasty may well be an effective and even cost-effective method for treating low back pain. If the technique is as good as its promoters suggest, then it should be straightforward to demonstrate its efficacy in a well-designed, controlled trial. Whereas reports such as the one by O'Brien et al add to our knowledge of how vertebroplasty can be performed, such articles cannot address if and when vertebroplasty should be done. The time is right to demonstrate the technique's advantages and convince the scientific community, as well as the public, of its worth.

JEFFREY G. JARVIK, M.D., M.P.H.
RICHARD A. DEYO, M.D., M.P.H.
University of Washington, Seattle, WA

References

1. World Medical Association.**Declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects [see comments].** *JAMA* 1997;277:925–926
2. Starr P. ***The Social Transformation of American Medicine.***New York: Basic Books; 1982
3. Osler W. ***Aequanimitas: With Other Addresses to Medical Students, Nurses and Practitioners of Medicine.***Philadelphia: Blakiston; 1905
4. Gordon R. ***The Alarming History of Medicine: Amusing Anecdotes from Hippocrates to Heart Transplants.***New York: St. Martin's Press; 1993
5. Moore TJ. ***Deadly Medicine: Why Tens of Thousands of Heart Patients Died in America's Worst Drug Disaster.***New York: Simon & Schuster; 1995
6. Fleming TR, DeMets DL. **Surrogate end points in clinical trials: are we being misled? [see comments].** *Ann Intern Med* 1996; 125:605–613
7. Pocock SJ. **Current issues in the design and interpretation of clinical trials.** *Br Med J (Clin Res Ed)* 1985;290:39–42
8. Turner JA, Deyo RA, Loeser JD, Von Korff M, Fordyce WE. **The importance of placebo effects in pain treatment and research [see comments].** *Jama* 1994;271:1609–1614
9. Dixon AS. **The evolution of clinical policies.** *Med Care* 1990; 28:201–220

Is Quality Medical Care Dead or Just Buried Alive?

Independent physician associations (IPAs) were thought to be one of the saviors of physicians during healthcare reform in the 1990s. These organizations formed to give physicians a greater say in the delivery of healthcare, including protecting the patients' interests and improving quality of care, while still implementing cost-effective strategies. In the year 2000, however, most IPAs are under severe economic pressure or becoming insolvent. How has this occurred, and how will this effect where we, as neuroradiologists, will be going in the future?

The 1980s brought increased pressure from healthcare buyers and consumers to decrease cost because of the perceived uncontrolled escalation in healthcare cost. Insurance companies quickly realized in the late 1980s that they could not shift this cost escalation to the consumers, and began to devise strategies to shift the risk and increased cost to hospitals and physicians. At this time, physicians perceived that they were losing control over patient-care decisions to insurers and hospitals, and were experiencing a continual decrease in reimbursement. Thus, both physicians and insurers supported the concept of forming physician organizations that could contract with hospitals and insurers, accepting some of the economic risk and gaining potential reward from the insurers. These new IPAs would contract with insurers, and if the IPAs were financially successful, physicians would benefit from increased reimbursement. The insurers saw an opportunity to shift financial and medical decision-making risk to the physicians by making the physician the ''gatekeeper.''

These IPAs initially used Kaiser Permanente as a model. In the late 1980s and early 1990s, for-profit health maintenance organizations (HMOs) formed or evolved from nonprofit companies because of the potential for large economic gain to the owners. This conversion, along with the development of for-profit companies to manage IPAs, created the environment of changing ''managed care'' into ''managed competition.''

In reality, this conversion led to the current problems with managed healthcare. First, HMOs usually marketed the same group of physicians and services to the consumers. Thus, the only opportunity to compete between the HMOs was on price. Second, a goal of ''for-profit'' HMOs was to maximize shareholder return. An HMO that broke even, but delivered excellent care, was considered a failure. Third, insurers and some physicians saw an opportunity to capitalize economically on the growing conflict between primary care and specialty medicine. This occurred by pitting the two groups against each other, fighting over a declining dollar. Fourth, these changes were occurring when the public financial market, many entrepreneurs, and physicians believed that healthcare companies were the ''new darlings of Wall Street.'' The stock price and value of these start-up companies (management companies and for-profit HMOs) were overvalued by standard evaluation methods.

The driving force in managed care and IPAs became cost containment, not cost effectiveness and quality of care improvement. The issue of quality care took a back seat to the market value of these companies. Primary care–controlled IPAs were the ''hot'' new way to control excessive specialty spending. Many primary care physicians were convinced that the road to economic parity was to become a manager of healthcare, not a deliverer. The rise in healthcare costs declined in the mid-1990s, and the publicly traded physician management companies boomed on Wall Street. The initial success of many of these IPAs was achieved through controlling Medicare hospitalization. Most IPAs assumed this could easily translate to the commercially insured business. Unfortunately, the ''great savings'' from controlling overhospitalization was not present in the commercial market. Additionally, most management companies did not have the technical ability to manage a large patient population and multiple physicians. Most HMOs, management companies, and IPAs fell into the trap of trying to pit the primary care physicians against the specialists as the major mechanism of lowering cost.

The main mechanism for increasing physician reimbursement was for IPAs to accept the risk of managing healthcare. Physicians assumed that they could better manage the economic risk of delivering healthcare than could the insurers. This was not based on past data, experience, or any factual information. Unfortunately, the insurers knew better and encouraged physicians to become the economic advocate and not the patient advocate. This new healthcare delivery system of for-profit HMOs, economically driven IPAs and publicly traded physician management companies, took only 5 years to implode. Consequently, many IPAs are facing bankruptcy. There is great consumer dissatisfaction with HMOs. Many consumers are wondering who is the patient advocate. These past 5 years have shown that the following assumptions are inadequate for quality patient care:

- Primary care physicians can manage costs by delaying or preventing referrals to specialists.
- IPAs can better manage risk than insurers.
- Specialty overuse is the primary cause of IPA failure.
- The correct treatment for good or excellent healthcare is known. It is difficult to determine the correct treatment because technology is evolving so rapidly (eg, positron emission tomography scanning, functional imaging, and diffusion imaging for the diagnosis of stroke). These rapid changes retarded the ability to determine how to deliver the best care for the lowest cost. Thus, insurers and

consumers reverted back to the goal of simply lowering cost. Unfortunately, physicians contributed to this process because they were reluctant to support the evidence-based algorithmic approach to medicine unless they developed the algorithm.

• IPAs should develop patient care algorithms and, thus, lower cost and remain the patient advocate.

The basic question of how to deliver cost-effective medicine (good healthcare at the lowest price) still has not been adequately addressed. Quality is like pornography, everyone knows it when they see it, but few can define it. What does the future hold for radiology, particularly neuroradiology, in this healthcare environment? If the past is any indication, it will not be long before there are rising healthcare costs again and, eventually, the industry and consumers will demand change. If there are no indicators by which to measure quality care, or a process to deliver ''better, cost-effective care,'' then the cycle will recur with the insurers and industry demanding and finding ways to decrease costs by decreasing reimbursement and withholding care. There will be few other choices. Thus, we are at a crossroads.

During the next several years there is an opportunity in this fee-for-service environment to develop vehicles to demonstrate quality care and cost-effectiveness. The American Society of Neuroradiology has an opportunity to be a leader in this endeavor. There are several issues we should address to enhance what we are doing. First, we should continue to evaluate new equipment with the emphasis on its cost-effectiveness for improved patient outcome, not just by whether its use will be reimbursable. Second, we should continue to develop imaging algorithms for diseases (eg, acute stroke, degenerative brain disease, carotid artery disease), and work with other specialties and insurers to determine what studies should be reimbursed and on what type of equipment. Third, we should work with other specialties to enhance patient care and shorten hospitalization time (for patients with spinal cord and head injuries, back disorders, or dementia, for example).

By being proactive and defining quality care and cost-effectiveness, we hopefully will not be driven back into the cycle of having patient care withheld, physician reimbursement decreased, or both to manage the cost of healthcare.

JAMES N. DREISBACH, M.D.
Member, Editorial Board