Myelopathy Associated with Cervical Spondylosis: A Frequently Unrecognized Disease

Donald I. Peterson, MD
Lloyd A. Dayes, MD
Loma Linda, California

Cervical spondylosis or chronic diskogenic disease of the cervical spine is a relatively common cause of myelopathy, but it is often not recognized or is incorrectly diagnosed. The clinical presentation may mimic several types of neurological disease including multiple sclerosis and amyotrophic lateral sclerosis. Even more frequently, and especially early in the course of the disease, neurologic impairment is not recognized and the symptoms are thought to be due to osteoarthritis. Early recognition of this condition is important since adequate treatment can prevent slowly progressive neurologic impairment. Knowledge of the pathophysiology of myelopathy due to cervical spondylosis and adequate radiographic evaluation will often lead to treatment that can prevent progressive spinal cord damage. Cervical spondylosis with myelopathy is one of the most frequently unrecognized and misdiagnosed, yet treatable, conditions affecting the nervous system.

Cervical spondylosis is a degenerative disease of the cervical spine which causes myelopathy if the degenerative process injures the spinal cord, or radiculopathy, if nerve roots are damaged in the spinal canal or as they leave the spinal canal. When this condition presents with symptoms due to nerve root impairment, the correct diagnosis is usually suspected, but when cervical spine disease presents chiefly as a myelopathy, it is often unrecognized or incorrectly diagnosed. For this reason, effective treatment may be delayed, resulting in progressive, irreversible damage to the spinal cord. This paper describes the mechanism of cord damage that may result from degenerative disease of the cervical spine and discusses the most important symptoms, signs, and radiographic criteria for diagnosis of myelopathy due to cervical spondylosis.

Although spinal cord damage from bone and soft tissue abnormality in the cervical area has been recognized since antiquity,[1] an awareness of the frequency with which this condition causes neurological abnormality with an understanding of its pathology and the radiographic criteria for diagnosis is relatively recent. The first complete description of the neurological abnormalities due to cervical spondylosis was by Brain et al.[2] Several other investigators have contributed to an increased understanding of this condition.[3-6]

From the Neurology Section and the Section of Neurosurgery, Loma Linda University, School of Medicine, Loma Linda, California. Requests for reprints should be addressed to Dr. Donald I. Peterson, Neurology Section, Loma Linda University, School of Medicine, Loma Linda, Calif 92354.

Etiology and Predisposing Factors

The etiology of the degenerative bony changes in cervical spondylosis that are associated with intervertebral diskogenic disease appears to be similar to that of osteoarthritis in other parts of the body and the incidence increases with age. Recent or remote head and neck trauma may be important contributory factors because they tend to increase the rate of progression of the pathological process. Congenital abnormalities of the cervical spine increase the incidence of degenerative changes. If congenital fusion of vertebrae is present, the most prominent changes occur at the intervertebral spaces adjacent to the fused vertebrae. Some degree of cervical spondylosis can be demonstrated radiographically in many people past the age of 40.[7,8] However, it usually is not associated with spinal cord pathology because there is adequate room for the cord in the spinal canal, even in the presence of prominent degenerative change, but myelopathy is especially likely to occur in patients with small spinal canals.

Pathophysiology

Cervical spondylosis consists of degenerative changes in the intervertebral disks and vertebrae associated with changes in adjacent soft tissue (Figure 1). As the disk degenerates it becomes thinner and this causes some buckling and posterior bulging of the annulus fibrosis and the posterior spinal ligament. There are bony changes in the margins of the vertebral bodies adjacent to the degenerating disk. These changes consist chiefly of irregular bone proliferation causing formation of osteophytes, which may project dorsolaterally into the vertebral foramina, or posteriorly into the spinal canal. These osseous changes along with the posterior bulging of the annulus fibrosus and posterior spinal ligament produce osseofibrous ridges, sometimes referred to as transverse bar formation when seen on posterior-anterior views by myelography. These projections into the spinal canal may decrease its diameter sufficiently to damage the spinal cord by compression or by interfering with its blood supply. These changes are usually greatest at the 4th, 5th, and 6th interspaces where the cervical spine is

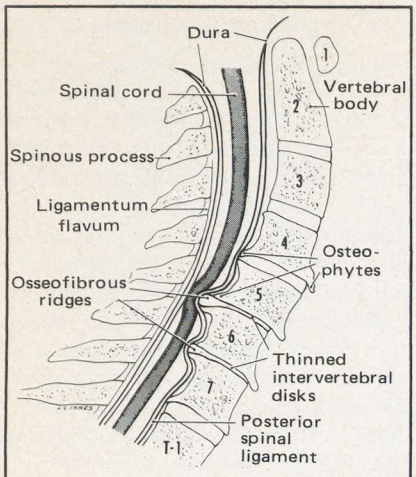

Figure 1. Midsagittal section of cervical spine and cord illustrating degenerative bone and disk changes causing spinal cord compression at the $C_{5,6}$ interspace. There is thinning of the disk with bulging of the annulus fibrosis, osteophyte formation, and vertebral subluxation causing decrease of spinal cord space.

most mobile during flexion and extension. As the disk deteriorates there is decrease in height of the disk space with some associated loss of joint stability, and this often results in malalignment of the vertebral bodies, thus further decreasing the space available for the spinal cord. During flexion and extension there may be considerable anterior-posterior displacement (subluxation) of one vertebra on another. These changes cause some thickening and indentation of the dura and there may be adhesions between the dura and the arachnoid.

Myelopathy may be exaggerated by compromise of the radicular arteries by foraminal stenosis produced by bony proliferation. This could render the cord marginally ischemic.

The spinal cord within its meningeal covering is well protected by the bony spinal canal. The average transverse diameter of the canal of the inter-pedicular distance in the cervical area is 30 to 34 mm.[9] Compromise of this diameter sufficiently to damage the spinal cord rarely occurs but shortening of the anterior-posterior or sagittal diameter is not infrequent and is often a predisposing cause of cervical spondylotic myelopathy because in canals with a small sagittal dimension osteophyte formation and soft tissue hypertrophy are more likely to compress the cord or interfere with its blood supply. Wolf et al[3] measured the anterior-posterior diameter of the cervical canal in 200 adults on films taken with a 72-inch tube to film distance. The average sagittal diameter at levels of 4th, 5th, 6th, and 7th cervical vertebrae was 17 mm (range 12 to 22 mm) when measured from the most posterior portion of the vertebral body to the most anterior portion of the spinous process at the above levels.

Payne and Spillane[4] reported that in patients with cervical spondylosis with myelopathy the canal diameter is smaller and averaged slightly greater than 14 mm at the midpoint of the body of the 6th cervical vertebra, so less space is available for the spinal cord. This measurement did not take into account the decrease of diameter of the spinal canal due to osseofibrous ridges at the intervertebral spaces, which further compromises the space for the spinal cord. Wolf et al[3] suggest a sagittal diameter of the spinal canal as projected on plain cervical spine films of 10 mm or less at any point is likely to be associated with cord damage, whereas a minimum diameter of 13 mm or greater suggests that there is adequate space for the spinal cord. These measurements (Figure 2) are helpful in determining which patients should have myelography but are not completely diagnostic in themselves because there may be considerable encroachment on the space available for the spinal cord by soft tissue not visible on plain radiograms (Figure 2). For this reason the cord space may be compromised even when the plain films suggest that it is adequate. In those patients with a spinal canal of short anterior-posterior diameter, the decrease in size appears to be due to shortened pedicles related to congenital and growth factors which in most cases are not well understood and are usually unassociated with other abnormalities. Decreased size of the spinal canal is often found in association with some dysplasias such as achondroplasia,[10] mucopolysaccharidosis[11,12] and fibrous dysplasia.[7]

Failure to recognize the importance of the sagittal diameter of the cervical spinal canal and its relation to the space required for the spinal cord is largely responsible for lack of recognition of cervical spondylosis as a cause of cord damage. This has apparently led to the conclusion by some investigators that cervical spondylosis is a relatively benign condition because in some cases there is no myelopathy, even though severe degenerative bone and disk changes are seen.[13] Minor degenerative changes in a small spinal canal may cause severe myelopathy, whereas severe bone and disk changes in a large canal may be relatively asymptomatic.

Symptoms and Signs

Symptoms and signs of cervical spondylosis with myelopathy are usually those of upper motor neuron impairment of the lower extremities. These consist of varying degrees of stiffness, slowness, clumsiness, and incoordination of movement associated with hyperreflexia, spasticity, and pathological reflexes. Sphincter impairment may occur in patients with more severe cord damage. Sensory symptoms and signs may be present but these are usually relatively minor compared with the motor system abnormalities. Decrease of vibration and position sense may occur. Impairment of touch and pain sensation are less frequently found. The upper extremities are often involved, but to a lesser degree than the lower extremities. Symptoms in the upper extremities usually consist of some clumsiness and slowness of movements associated with hyperactive reflexes. Tremor and mild ataxia may be present. There may also be considerable loss of strength and muscle bulk. If the cervical nerve roots are involved, pain and paresthesias often occur in the upper extremities and in some instances decrease of sensation in the area supplied by the involved nerve root. Often there is weakness and decrease of deep tendon reflexes due to lower motor neuron impairment at the level of the cervical cord and root pathology. In these cases there may be a combination of upper and lower motor neuron abnormalities in the upper extremities. Neck and shoulder pain and headache are frequently associated symptoms when radiculopathy is also present.[14] The headache is chiefly in the occipital region but may also spread to other parts of the head. The headaches are often diagnosed incorrectly as "tension headaches" or when they are unilateral they are sometimes called "occipital migraine." The pains in the

shoulder, medial scapular area and upper arm are thought by Cloward[15] to be due to referred pain secondary to involvement of structures supplied by the sinuvertebral nerves which transmit pain sensation from the intervertebral disks.

Unusual and difficult to explain symptoms sometimes occur. Cervical spondylosis with myelopathy, as well as cervical cord disease due to other etiologies, may rarely cause pain in the lower extremities and be confused with lumbar disk disease, especially if asymptomatic lumbar diskogenic disease is demonstrated radiographically. With proper clinical evaluation, however, they can be differentiated. As pointed out by Langfitt et al,[16] pain in the legs from cervical cord disease is of a different quality than pain due to nerve root disease in the lumbosacral area. It has more of a deep, burning, boring, and poorly localized quality than radicular pain, and it is not associated with other symptoms and findings of nerve root involvement in the lumbosacral region. The cause of the pain is not known but it may be related to spasticity in the lower extremities due to upper motor neuron impairment. Other symptoms that are sometimes associated with cervical spondylosis, but for which the cause is poorly understood, are dizziness and true vertigo and occasionally tinnitus. When these symptoms are associated with cervical spondylosis they are often relieved by surgical treatment of the cervical spine disease, and sometimes they are temporarily relieved after greater occipital nerve block, when this is done for symptomatic relief of the associated occipital headache.

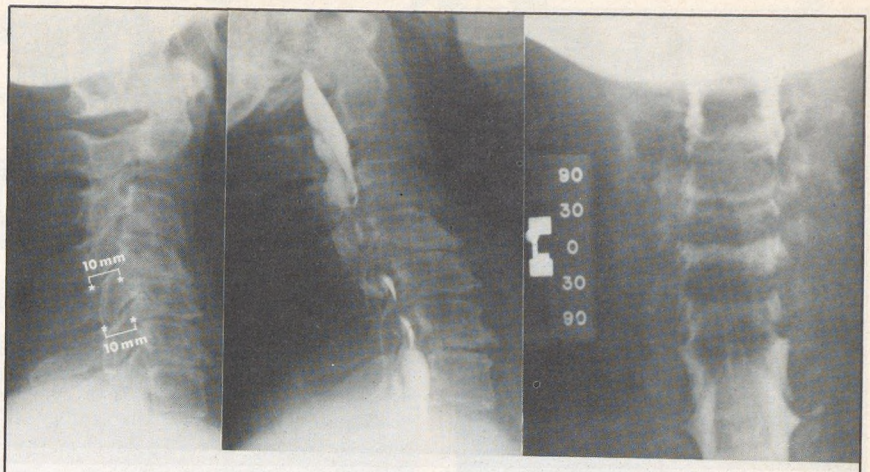

Figure 2. Plain x-ray and myelogram views of the cervical spine of a patient with myelopathy. The cord space is decreased to 10 mm on the plain film. On the myelogram contrast material does not flow between the cord and osseofibrous ridges.

Radiographic Examination

When disease affecting the spinal cord in the cervical area is suspected after neurological evaluation, radiological examination is very helpful in determining whether or not it is caused by cervical spondylosis. In any case where there is less than 11 mm of space available for the spinal cord when measured at the site of greatest narrowing due to osseofibrous ridge or osteophyte formation on lateral cervical spine films, myelopathy is likely to be present. Cervical myelography, by either positive contrast or air techniques, helps confirm the diagnosis and is especially helpful in cases with involvement at multiple levels because myelography helps identify the levels where the cord is most likely to be compressed. It cannot be too strongly emphasized that the severity of the degenerative changes, including osteophyte formation, spurring, bridging, thinning of the disk, subluxation, and "hypertrophy of the ligamentum flavum" as seen on plain radiograms and by myelography, is of much less importance than the space available for the spinal cord, since it is cord damage and not bone and disk degeneration per se which produces the impairment of the function due to cervical spondylosis with myelopathy. The measurements of the sagittal diameter of the spinal canal referred to above apply to plain films taken at 72 inches and not to lateral myelogram views taken at a lesser distance. Figures 2 and 3 illustrate radiographic changes of cervical spondylosis.

Although the importance of the sagittal dimension of the canal and cord space has been emphasized by several investigators,[3,5,8] it is still not generally recognized and commonly no reference is made to this important parameter of cervical spine radiography in reports of cervical spine x-rays. Inadequate interpretation of x-rays of the cervical spine is a major factor in failure to recognize myelopathy due to cervical spondylosis.

Treatment

Treatment of cervical spondylosis can be either medical or surgical. Medical treatment consists in general of immobilization by use of a cervical collar to prevent trauma and to hold the head in a neutral position. Bradshaw[17] reported that some patients with myelopathy improved or remained unchanged when treated with a neck collar, but many others deteriorated after 6 to 18 months. The majority of those with spondylosis at two or more levels did badly in that the impairment of function in their legs progressed, but more than half of those with involvement at one level did not get worse and some improved. Other methods of treatment include analgesics and antispasmodics for symptomatic relief of pain and to decrease muscle spasm. Traction, muscle strengthening exercises, and maintenance of proper posture may in some cases decrease the rate of degenerative changes. These methods are inadequate if there is spinal cord damage.

More effective treatment consists of increasing the space available for the spinal cord by surgery. Surgery should be considered, especially in those cases with early myelopathy, for by this means irreversible damage to the cord may be prevented. Surgical therapy consists of either laminectomy by the posterior approach, anterior interbody

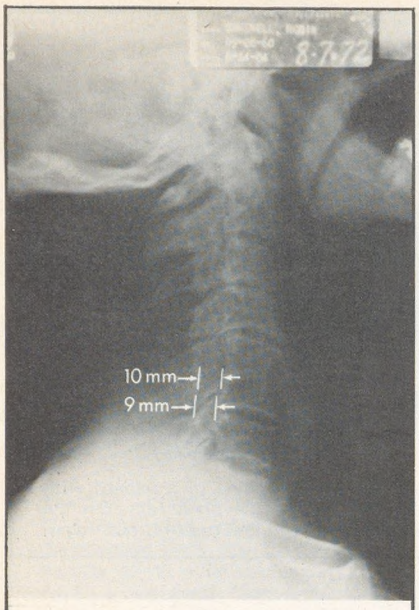

Figure 3. Lateral cervical spine x-ray of a patient who has been incorrectly diagnosed as multiple sclerosis for ten years. Degenerative changes are mild but the sagittal diameter of the spinal canal is very small. Congenital fusion of $C_{5,6}$ is present.

fusion by the method of Smith and Robinson[18] or anterior diskectomy and interbody fusion by the method of Cloward.[19] To a large extent posterior surgical approach by laminectomy has given place to the anterior approach with diskectomy and fusion. Although the type of surgical procedure depends somewhat on the experience of the operator and the presentation of the patient, good results are usually obtained by anterior diskectomy and interbody fusion,[5,15] and this appears to be the treatment of choice. Most effective management in each case can be best determined after neurosurgical consultation with consideration given to the degree of impairment, rate or progression of disease, age of the patient, and other variables.

Prognosis

Reports in the literature regarding prognosis of cervical spondylosis reveal considerable difference in opinion. Lees and Turner[13] reported a study of the natural history of cervical spondylosis in 44 patients with myelopathy and 51 patients with symptoms referable to the neck, shoulders, arms, and hands which were described as cervical spondylosis without myelopathy. These investigators found that nonprogressive disability was the rule in patients with myelopathy and none of the patients of the non-myelopathy group developed any myelopathic symptoms during the period of observation. This report does not include any reference to the dimensions of the cervical canal of the patients.

If the spinal canals of the patients in this series were adequate, there was little risk of myelopathy, even with severe bone and disk changes. Many investigators found that the course of cervical spondylosis with myelopathy is much less benign.[2,5,17,20] It is not unusual to find patients with slowly progressive deterioration to complete incapacity due to loss of function of the lower extremities, sometimes associated with sphincter loss, from myelopathy due to cervical spondylosis that could have been effectively treated surgically if it had been recognized earlier. Prognosis is usually adversely affected by increased age, advanced degenerative changes, trauma, and long duration of symptoms. Some of these factors, however, are relatively unimportant in themselves as causes of myelopathy unless the patient has a small spinal canal.

Persons with mild myelopathy or those who have no symptoms but have marginal cord space may suffer severe damage from acute trauma; especially hyperextension injury. This type of injury may cause sudden cord compression resulting in quadriparesis, sphincter impairment and sensory loss below the level of the lesion, even without fracture or dislocation of vertebrae.

Prognosis for cervical spondylosis with myelopathy when treated surgically by anterior diskectomy and interbody fusion is usually good in that progression of the myelopathy is often stopped and in many cases there is some improvement, especially in decrease of spasticity and pain.[5] Severe cord damage, if present, cannot be expected to improve significantly, and for this reason early diagnosis and treatment are important.

Summary

Although degenerative disease of the cervical spine as a cause of myelopathy has been known for many years, its frequency and clinical presentation and the criteria for radiographic diagnosis are often not recognized. There is some controversy regarding most effective management but early diagnosis by proper neurological evaluation of the patient with this condition usually leads to treatment that can prevent progressive impairment of function. Cervical spondylosis with myelopathy is one of the most frequently unrecognized and misdiagnosed, yet treatable, conditions affecting the nervous system.

References

1. Wilkinson M: Cervical Spondylosis, ed 2, Philadelphia, WB Saunders, 1971
2. Brain WR, Northfield D, Wilkinson M: The neurological manifestations of cervical spondylosis. Brain 75:187-255, 1952
3. Wolf BS, Khilnani M, Malis L: The sagittal diameter of the bony cervical spinal canal and its significance in cervical spondylosis. J Mt Sinai Hosp 23:283-292, 1956
4. Payne EE, Spillane JD: The cervical spine, and anatomico-pathological study of 70 specimens (using a special technique) with particular reference to the problem of cervical spondylosis. Brain 80:571-593, 1964
5. Crandall PH, Batzdorf U: Cervical spondylotic myelopathy. J Neurosurg 25:57-66, 1966
6. Epstein JA, Carras R, Epstein BS, et al: Myelopathy in cervical spondylosis with vertebral subluxation and hyperlordosis. J Neurosurg 32:421-426, 1970
7. McRae DL: The significance of abnormalities of the cervical spine. Amer J Roentgenol Radium Ther Nucl Med 84:3-25, 1960
8. Pallis C, Jones AM, Spillane JD: Cervical spondylosis, incidence and implication. Brain 77:274-289, 1954
9. Elsberg CA, Dyke CG: The diagnosis and localization of tumors of the spinal cord by means of measurements made on the x-ray films of the vertebrae, and the correlation of clinical and x-ray findings. Bull Neurol Inst NY 3:359-393, 1934
10. Maroteaux P, Lamy M: Achondroplasia in man and animals. Clin Orthop 33:91-103, 1964
11. Einhorn NH, Moore JR, Rountree LG: Osteochondrodystrophia deformans (Morquio's disease). Amer J Dis Child 72:536-544, 1946
12. Peterson D, Bacchus H, Seaich L, et al: Myelopathy associated with Marteaux-Lamy syndrome. Arch Neurol 32:127-129, 1975
13. Lees F, Turner JQ: Natural history and prognosis of cervical spondylosis. Br Med J 2:1607-1610, 1963
14. Peterson D, Austin G, Dayes L: Headaches associated with diskogenic disease of the cervical spine. Bull Los Angeles Neurol Soc 40:96-100, 1975
15. Cloward RB: The clinical significance of the sinu-vertebral nerve of the cervical spine in relation to the cervical disk syndrome. J Neurol Neurosurg Psychiatry 23:321-326, 1960
16. Langfitt TW, Elliott FA: Pain in the back and legs caused by cervical spinal cord compression. JAMA 200:382-385, 1967
17. Bradshaw P: Some aspects of cervical spondylosis. Q J Med 26:177-208, 1957
18. Robinson R, Smith GW: Anterolateral cervical disc removal and interbody fusion for cervical disc syndrome. Bull Johns Hopkins Hosp 96:223-224, 1955
19. Cloward RB: The anterior approach for removal of ruptured cervical disks. J Neurosurg 15:602-617, 1958
20. Spillane JD, Lloyd GHT: The diagnosis of lesions of the spinal cord in association with "osteoarthritic" disease of the cervical spine. Brain 75:177-186, 1952

Continued on page 238

angiography or the measurement of the extent of coronary atherosclerosis among patients who have died subsequent to entry into the study.

Experimental Studies

Descriptive and analytic studies do not include modification of the independent variables such as nutrient intake, serum cholesterol levels, etc. Experimental approaches, depend on a modification of such variables and the subsequent determination of changes in the dependent variable, ie, the incidence of disease or mortality from heart disease. These studies may be of two major types:

1. Natural experiments
2. Planned or clinical trials

The best examples of natural experiments are migrant studies in which a group of people from either a high- or low-risk-factor disease area migrate to an area which has markedly different environmental or social characteristics and incidence of disease. For example, Japanese migrants to Hawaii and California were compared with Japanese still living in Japan and the non-Japanese population in Hawaii and California. Dietary changes were related to the serum cholesterol levels and subsequent disease incidence. Such migrant studies have several basic flaws. First, migrants are self-selected and not representative of the country of origin. Second, migration may occur after the critical point of disease development. Thus, studies of second generation migrants may be required. Third, migrants do not necessarily adopt the same habits as the host population. Conversely, numerous changes occur and may confound the relationship between, let us say, nutrient and serum cholesterol levels.

A second type of natural experiment occurs when an event results in a major change in the behavior of a population over time and the changes in incidence or mortality of disease can be measured during this time period or following the event. A classical example of such events include the influence of nutritional deprivation during World War II on the incidence of disease and extent of coronary atherosclerosis. Another example, aside from the cardiovascular area, was the effect of the implementation of alcohol prohibition in the United States on death rates due to cirrhosis of the liver. Since such events are often unplanned, collection of data before, during, and after the event are usually inadequate for meaningful analysis. Sometimes, however, such events may be the only way of specifically testing a hypothesis.

Planned experimental studies or clinical trials have been described in detail. Nutritional studies that might be classified as planned or experimental trials have included the deliberate manipulation of specific nutrients in order to measure the changes in blood lipid and lipoprotein levels, and modification of blood lipid levels by either diet or drugs in order to measure variations in incidence of disease. The earliest studies would be classified as sequential time trials. For example, a group of volunteers would have specific nutrients modified and changes in blood lipid or lipoprotein levels before and after the dietary changes compared. These studies were beset by the difficulties of controlling nutrient intake in a "free-living" population as well as the probable changes in other variables. Studies in which participants were admitted to controlled metabolic research units partially alleviated these problems.

A second major problem, however, has been a phenomenon known as regression to the mean: the tendency of extreme values of a variable or biologic characteristic to regress naturally toward the mean levels over time, independent of any manipulation. Thus, investigators who evaluate the reduction of high serum cholesterol levels by nutrient changes cannot be sure whether the reduction in serum cholesterol is due to the experimental manipulation or regression to the mean unless a selected control group is available. A similar phenomenon could be observed when individuals with a low serum cholesterol level were observed and a change in dietary intake, ie, more fat in the diet, was determined. It became apparent that the selection of a suitable control group as a comparison for the experimental group was critical to any of these dietary manipulation studies, since both the experimental and control groups would be exposed equally to the phenomenon of regression to the mean and any differences between the experimental and control group

Continued from page 241

Continued from page 238

could be attributed primarily to the experimental manipulation of the diet.

Another problem was the temporal trend in quality control within the laboratories. It became apparent that there was a need for quality control techniques within the laboratories, including duplicate samples, and saving aliquot samples for repeat measurements later in order to avoid the confounding effects of the changes in laboratory techniques over time.

Blinded studies, especially those in which neither the participant nor the investigator knew the status of the experimental or control group, minimized the potential influences of extraneous variables on test results. One would expect that such influences would be relatively evenly distributed between experimental and control groups at entry to the study if the experimental and control groups had originally been randomly allocated. Although blinding is fairly successful in studies involving medications, nutritional studies have presented special problems.

The National Diet Heart Study was a double blind approach to nutrient intake and effects on blood lipid levels. The blinding of this study and other nutritional studies requires that the nutrients be supplied to the participants in such a way that neither the subject nor investigators are able to identify the experimental and control nutritional changes. The preparation of the food and its distribution are extremely costly. Blinding of the nutritional studies is somewhat easier when the studies are done in so-called captive populations that are all eating most of their meals at a specific site. For example, studies have been done on patients in mental hospitals, domiciliary residents of the Veteran's Administration Hospital, and on military and prison personnel. Although blinding of the diet is easier in some of these populations, the lack of representativeness in such samples is quite apparent.

An alternate approach to randomizing individuals into experimental and control groups has been to randomize the population groups. A classic example of such an approach has been the Finnish mental hospital studies of dietary change, serum cholesteral levels, and mortality and morbidity due to heart attack. The diet in one mental hospital was manipulated and the other hospital was used as a control. Later on in the study a cross-over was accomplished so that the experimental hospital became the control hospital and vice versa. This approach is obviously much more economical than randomizing each individual to either the experimental or control group. However, the power of this study in demonstrating a difference between the two groups is greatly weakened. The sample size really consists of just two or three institutions rather than several thousand people within each institution, and only differences between the two institutions are meaningful for analysis.

It is another giant step forward from experimental studies that attempt to show the effects of dietary modification on serum cholesterol levels and those studies which attempt to show the effect of change of serum cholesterol on the risk of heart attack within the population. Clinical trials that have been designed to demonstrate a reduction of mortality or incidence of coronary heart disease because of a change in blood lipid levels have generally been part of the previously described studies of changes in blood lipid levels. The selection of the characteristics for participants in such studies and determination of specific end points have been most interesting problems for epidemiologists.

The studies can be conceivably divided into primary and secondary prevention. Primary prevention studies are those in which the subject has no clinical evidence of disease prior to entry into the study. Secondary prevention studies involve subjects with preexisting clinical heart disease. It should be noted that these studies classify individuals into those free of disease on the basis of clinical heart disease and not on the extent of the underlying coronary atherosclerosis.

The studies involving those individuals who already have had a heart attack probably include mainly subjects who presently have extensive existing coronary atherosclerosis of long-standing duration. The proposed hypothesis is that reduction of serum lipid levels will prevent further exten-

Continued on page 243

Continued from page 241

sion of the atherosclerosis, possibly reduction in the extent of coronary artery stenosis, or influence other secondary processes such as thrombosis and embolism. The likelihood of recurrent heart attacks and subsequent mortality is reduced. A few studies of secondary prevention have generally not shown a reduction of risk of heart attack or death following the modest reduction of serum cholesterol levels, usually about 10 percent change (in the serum cholesterol level), and thus conclude that changes in the level of serum cholesterol are probably ineffective in reducing subsequent mortality and morbidity among individuals who have already had an initial heart attack and presumably extensive coronary atherosclerosis.

Very few primary prevention trials have been completed. Many important issues have been raised about such trials, as for instance, should the subjects be relatively young and essentially free of coronary atherosclerosis? The number of subjects required for such a trial among young people would be very large because of the small number of events – ie, heart attacks or death – that occur over time. Should the studies be limited to dietary intervention only, or include other risk factors such as smoking and elevated blood pressure – that is, multiple risk factor trial? Studies limited to nutritional changes only have an inherent bias since the treated group may modify their risk factors more than the so-called control group. For example, weight reduction as part of a dietary manipulation in the treated group may not only lower the serum cholesterol and triglyceride levels, but may also lower blood pressure. More subjects in the treated group may become enthusiastic about the preventive medical aspects of the program, give up cigarette smoking or seek medical care for minor symptomatology, and thus, perhaps prevent subsequent heart attack. Would it be ethical in such trials to encourage individuals to continue to smoke and maintain high blood pressure? Blinding the nutritional intervention might reduce the probability of variations in the other risk factors among both cases and so-called experimental groups, but as previously described, blinding of such trials is inordinately expensive and difficult to accomplish.

What should the end point of these studies be? Total deaths, deaths due to arteriosclerotic heart disease without noting a change in total mortality? Could changes in the extent of coronary atherosclerosis or peripheral atherosclerosis be used as an end point as has been suggested? Perhaps the most important question to be answered is whether dietary changes result in a decrease in atherosclerotic disease or at least prevent further evolution of disease as compared to so-called control groups. If this hypothesis could be tested and proven in man, then at least a relationship between nutrients and blood lipid levels and atherosclerosis could be more firmly established.

Would it be possible to lower the serum cholesterol level in an individual with drugs and note the effects on atherosclerosis and subsequent disease, and presume that if such changes occur with drugs similar changes would occur with nutritional changes if they also lowered serum cholesterol levels? It might be difficult to extrapolate the results from a drug trial to the expected results of nutritional changes which might result in the same effect on blood lipid levels, but not necessarily the same effect on the underlying atherosclerosis or subsequent disease.

Finally, can animal experimental models such as primates be used as a substitute for these difficult clinical trials? Would evidence of regression of different stages of atherosclerosis in primates following dietary manipulation be accepted as evidence that dietary manipulation would yield similar results in man? Would such evidence justify the assumption that the regression of atherosclerotic lesions would result in decreased risk of heart attack in man? Certainly most investigators would feel insecure without definitive evidence of change in man. Thus, clinical trials in man (experimental epidemiology) can probably not be replaced by laboratory experimental studies with animals. The experimental epidemiologic study of the relationship between nutrient intake and heart disease is not easy. The protagonist may accept the descriptive and analytical evidence of relationship between diet and atherosclerosis and the impressive animal experimental literature as proof enough, as public health and preventive medicine approaches reduction of the incidence or mortality of heart attack through dietary manipulation. However, the skeptic requires further, more detailed solutions. Many of the requested solutions to the problems may not be obtainable in man, at least if the ethical considerations of human experimentation are observed.

Bibliography

Blackburn H: Multifactor preventive trials (MFT) in coronary heart disease. In Stewart GT (ed): Trends in Epidemiology. Thomas, 1972, pp 212-230

Blackburn H: Progress in the epidemiology and prevention of coronary heart disease. In Yu PN, Goodwin JF (eds): Progress in Cardiology. Lea, Febiger, 1974, pp 1-35

Connor WE, Connor SL: The key role of nutritional factors in the prevention of coronary heart disease. Prev Med 1:49-83, 1972

Dayton S, Pearce ML: Prevention of coronary heart disease and other complications of atherosclerosis by modified diet. Am J Med 46:751-762, 1969

Epstein LM, Reshef A, Abramson JH, Bailik O: Validity of a short dietary questionnaire. Isr J Med Sci 6:589-597, 1970

Goldstein JL, Schrott HG, Hazzard WR, et al: Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. J Clin Invest 52:1544-1568, 1973

Goldstein JL, Brown MS: Familial hypercholesterolemia, a genetic regulatory defect in cholesterol metabolism. Am J Med 58:147-150, 1975

Hankin JH, Messinger HB, Stallones RA: A short dietary method for epidemiologic studies. IV. Evaluation of questionnaire. Am J Epidemiol 91: 562-567, 1970

Kahn HA: Change in serum cholesterol associated with changes in the United States civilian diet, 1909-1965. Am J Clin Nutr 23:879-882, 1970

Kannel WB, Castelli WP, Gordon T, McNamara PM: Serum cholesterol, lipoproteins, and the risk of coronary heart disease. Ann Intern Med 74:1-12, 1971

Kannel WB, Gordon T (eds): Framingham Study; An Epidemiological Investigation of Cardiovascular Disease. Section 24. The Framingham Diet Study. Diet and the regulation of serum cholesterol. April, 1970

Keys A: The diet and plasma lipids in the etiology of coronary heart disease. In Russek HI, Zohman BL (eds): Coronary Heart Disease. Philadelphia. Lippincott, 1971, pp 59-74

Mojonnier L, Hall Y: The national diet-heart study – assessment of dietary adherence. J Am Diet Assoc 52:288-292, 1968

Moore MC, Moore EM, Beasley CH, Hankins GJ, Judlin BC: Dietary-atherosclerosis study on deceased persons. J Am Diet Assoc 56:13-22, 1970

Primary Prevention of Atherosclerotic Disease. Report of inter-society commission for heart disease resources. Circulation XLII:A-55-95, 1970

Reshef A, Epstein LM: Reliability of a dietary questionnaire. Am J Clin Nutr 25:91-95, 1972

Rose G, Blackburn H, Keys A, et al: Colon cancer and blood cholesterol. Lancet 2, Feb 9, 1974, pp 181-183

Turner R, Ball K: Prevention of coronary heart disease, a counterblast to present inactivity. Lancet 2, Nov 17, 1973, pp 1137-1140

Turpeinen O: Primary prevention of coronary heart disease by diet. Ann Clin Res 3:433-438, 1971

Book Excerpts

The following article has been selected by the Publisher from its new book, *Legal Medicine 1976,* edited by Cyril H. Wecht, in the hope that it will have immediate usefulness to our readers who otherwise might not have had access to it.

PSRO and the Dissolution of the Malpractice Suit*

Henry E. Simmons, BS
John R. Ball, JD, MD

Introduction

Professional Standards Review Organizations (PSROs), established by the Social Security Amendments of 1972,[1] are intended to review the appropriateness and quality of services provided beneficiaries of federal health reimbursement mechanisms, popularly known as Medicare, Medicaid, and Maternal and Child Health Care.[2] This review of medical services is conducted by community organizations composed of physicians.[3] PSRO was originally intended to serve two functions: to act to reduce the cost of medical care and to act to improve its quality.

The Senate Finance Committee Report of the Amendments emphasizes the cost impact by beginning its discussion of the PSRO provision with a discourse on the increasing costs of care.[4] The Report notes the increase in unit costs of all forms of health-care services, and the increase in numbers of services provided beneficiaries of Titles V, XVIII, and XIX of the Social Security Act.[5] Coupled with the concern of the Committee on Finance over costs was a concern that a significant proportion of health services provided may not be medically necessary. The Report states, for example, that "Unnecessary hospitalization and unnecessary surgery are not consistent with proper health care."[6] It was in the context of this dual concern that PSRO was born.

This article will be primarily concerned with the quality control function of PSROs since it is that aspect which will have an impact on the present medical malpractice system. It is the contention of this article that Professional Standards Review Organizations have the potential for reducing malpractice incidents, patient injury based on those incidents, and legal action resulting from those injuries. In addition, the program is likely to decrease the incidence of defensive medicine – defined as "medically unjustified care provided ... for the purpose of reducing the possibility of a malpractice suit."[7] These results will flow initially from the increase in the quality of medical care services expected to occur as review of those services becomes general.[8] That is, as quality of care increases, the number of malpractice suits should decrease.[9] In addition, defensive medicine should decrease in incidence as physicians gain confidence in PSRO developed norms and standards of care and cease overutilization of diagnostic and treatment modalities previously engaged in because of the fear of suit.[10]

The civil immunity provision of the PSRO Act,[11] providing for immunity from civil liability for health-care providers who follow PSRO norms and standards, will be useful in both regards. As an incentive for physicians to comply with PSRO, it will aid in the review and enhancement of the quality of care; and, in providing immunity only if the standards are followed, it suggests that utilization beyond the standards is unnecessary from either a medical or a suit prevention viewpoint.[12]

The greatest impact of PSRO on malpractice, however, could be evident in the dissolution of the malpractice system. Presently, the possibility that a physician may be sued and held liable for patient injury is one of the major controls on the quality of outcomes of care.[13] Although the quantitative impact of that control is not known,[14] it is alleged to be highly significant. The providers of care allege that the possibility of suit causes defensive medicine which increases the costs of health services and is unnecessary for the medical care of the patient.[15] Many consumers of care argue that the threat of suit is the final (or in some cases the only) check on physician malpractice.[16] PSRO has the potential to answer, in principle, both arguments. By developing standards of care, the overutilization of medical services caused by defensive medicine should decrease. In addition, peer review with statutory controls can become more effective and efficient than the malpractice system[17] in controlling the quality of outcomes of care.[18]

The incidence of professional liability suits is related to, but is not a sole function of, the quality of medical care.[19] While PSRO will have the direct effect of decreasing malpractice suits based on poor quality of care, it will also lead indirectly to a decrease in the number of suits based on injury from other causes, such as the inherent risks of certain procedures. Designed in part to curb medically unnecessary care, and thus overutilization,[20] PSRO will affect primarily the incidence of injuries caused by errors of commission. However, because of its development of standards of care, it will also affect injury incidents resulting from errors of omission. Thus, the more important and more general impact PSRO will have on malpractice is the set of effects that the promulgation of norms, standards, and criteria will bring about.[21]

Another major justification of the present malpractice system is that it serves to compensate the injured patient. This article will not attempt to analyze the effect of PSRO on this aspect of malpractice other than to express the hope that once PSRO has created a new environment of quality control a more equitable and efficient system of patient compensation will be instituted. It is, therefore, possible that PSRO will initiate a long-term process that will entail the dissolution of the fault system in medical accident injury.

Continued on page 245

*The authors express their appreciation to the University of Toledo Law Review for the opportunity to present their views on this important topic, but would emphasize that the views expressed are solely theirs and are not necessarily those of the Office of Professional Standards Review nor of the Department of Health, Education and Welfare. From 6 UNIV. TOLEDO LAW REV. 739 (1975).

Continued from page 244

Control Mechanisms and the Quality of Care

Although there is little agreement concerning what "quality of health care" is, there is general agreement on how it can be measured conceptually. Controls over the quality of care can be categorized into three general time frames or measures: input, process, and outcomes.[22]

The control of the quality of input is applicable to facilities and to personnel.[23] Examples of facility input measures are facility licensure and accreditation and other governmental, agency, and public health standards. Measures of personnel input are educational standards, professional licensure, and specialty group certification. All are static measures, evaluating a state of knowledge or ability at a particular point in time, and as such, may have no bearing on the quality of medical care.[24]

Like input measures and outcomes measures, process measures may also be applicable to both facilities and personnel. Examples of process measures, or means of quality control relating to process, are peer review, professional memberships (which may also be measures of input control), licensure revocation procedures (which occasionally are outcomes measures), and utilization review. Process measures of quality control have more relation to the legislatively required activities of PSRO and to the quality of care than do measures of input.[25] Primarily, they are methods which are susceptible to the input of data. PSRO mandates the local setting of standards of practice with regard to certain norms and criteria.[26] That is, it requires the setting of standards related to the process and result of patient care based on objective data, and more related to the quality of care than input measures have been shown to be.

Examples of measures of the quality of the outcomes of care are less numerous than those of input or process. While some specific process elements have some relation to outcome, presently the only true outcome control measure is the malpractice suit.[27] Not only is the threat of suit alleged to stimulate defensive practices (and thus have some of the characteristics of process control) but the possibility of suit serves as the final check on medical care. The malpractice suit comes into play when both input and process controls have failed to prevent a medically related injury.

All measures of quality control must be subject to an examination of their efficiency in enhancing quality. However, the effectiveness of malpractice in insuring outcomes quality has never been critically examined. Perhaps one of the reasons for this lack of scientific data is that the malpractice system has other facets and purposes than just quality control. For example, this system serves as a means of compensation for the injured patient. However, it is unclear that the malpractice system as it presently exists does an efficient, just or adequate job of either outcomes control or patient compensation. The premises on which its existence is based should therefore be subject to scrutiny. What is needed are studies which determine both the total cost, including social, political, and ethical considerations of the malpractice system and the relative worth of malpractice as both a control on the quality of care[28] and as a compensatory mechanism. Then the conflicting interests and values involved must be balanced and alternative systems considered. However, until the total costs of the systems are known and until the total benefits from the standpoint of improved quality of care and just compensation of injured patients are determined, it will be impossible to objectively evaluate malpractice as a social system.[29]

PSRO as a Means of Quality Control

Neither the goals of PSRO (control of costs and improvement of quality) nor its method (peer review) are unique to the program. The former are concerns of providers, consumers, and payors of health care, and the latter has been utilized for some time in process measures of control. But PSRO is different from any previous health control system in several important ways. First, it legislatively requires review of services for which the federal government is a third party payor. Secondly, it contains a structure for that review based on norms, standards, and criteria of care. Thirdly, it builds in both incentives to comply and sanctions against noncompliance.

Legislatively Required Review

The review of services required by the legislation is delegated to "qualified organizations"[30] composed, in the first instance, of physicians. This delegation is based on the premise that physicians are best qualified to judge the necessity and usefulness of physicians' services.[31] Arguments against this type of regulation are that it allows the regulated industry to control the regulators,[32] that there is no empirical data to support the conclusion that physicians are best qualified to judge necessity and usefulness of medical service, and that there are insufficient standards set up by the legislation to support delegation to a non-governmental unit.[33] However, this type of self-regulation appears to be a valid initial premise in a field as specialized as medicine. Moreover, the review of both medical process and outcome of medical treatment which is dictated by PSRO legislation is a definite step forward from the present "outcome review" afforded by malpractice suits.[34]

Neither of the two primary parties to medical care services, the providers nor the consumers, have expressed total satisfaction with PSRO. Many of the providers consider legislatively required peer review to be overreaching on the part of government and an unwarranted interference in medical practice,[35] while consumers view the system as inadequate to assure the quality of care.[36] While those who drafted and passed the legislation and those responsible for its administration

Continued on page 246

Continued from page 245

understand the arguments on both sides, what should be emphasized is that however imperfect PSRO may be, it represents a significant step in the direction of quality assurance as well as cost control – a step the value of which has yet to be seen. The responsibility for setting standards of care and for reviewing that care remains with the providers,[37] whose role should require continuing reviews even in the absence of law. Thus, the providers, rather than attacking the system of PSRO from without, should avail themselves of the very real opportunity to mold it from within into an effective force for quality practice. The consumers, on the other hand, should view PSRO as at least a step toward assuring good outcomes, and as a step forward from a system in which the individual provider alone determines the necessity for utilization of services and facilities. For example, before the advent of PSRO, individual providers, in general determined what amounts would be paid out by Medicare and Medicaid.[38] With PSRO those decisions are subject to review and measurement by standards which will reflect not only quality care but also local need.

Structure of Review

The structure of the review process substantially revolves around norms, standards, and criteria. While norms and standards are mentioned specifically in the legislation,[39] they are not well-defined by it and it was not without some effort that the National Professional Standards Review Council and the Task Force on Guidelines of Care of the American Medical Association agreed on specific working definitions of the terms.[40] Norms, as utilized by the local PSRO, are independent, numerical descriptions of actual practice. An example might be that ten percent of a certain population have hypertension. Criteria are statements which to an extent define good practice. For example, a criterion as used by PSRO might be that hypertension is a discoverable, treatable disease. Standards based on that criterion would reflect variations in methods by which the disease could be diagnosed and treated. Thus, there are relations among the terms but they form more of a spectrum of care rather than a continuum.

The legislation further mandates that the Department of Health, Education and Welfare and the National Professional Standards Review Council provide technical assistance to local areas in the development of standards.[41] This provision assures that national resources will be available to the PSRO in its determination of standards. The 203 local review organizations will not need to separately determine all the elements – norms, standards, and criteria – on which care is to be based and reviewed, but may draw on knowledge gained more efficiently at a national scale. Thus, the structure of the review process generates a potential for cost savings. In addition, the structure of PSRO would appear to require the determination, on a national scale, of the medical procedures which benefit the consumer and those which do not. This in turn would require federal interest in and support for research to develop data concerning clinical efficacy and the cost/benefit ratio of clinical practices which would be of considerable value to local providers.

While it may be argued that both control of the information on which decisions as to standards are based and control of the dissemination of that information may lead to "cookbook medicine" based on national standards of care, the legislation and the administration of PSRO require only that the data generated at the national level be taken into account in local standards-making. Ultimate power and responsibility for standards-setting reside at the local level.[42] Is it true that if local PSROs desire to utilize nationally-produced data in standards-setting decisions, standards of care and presumably practice patterns might become more standardized, tending toward the development of a national uniform standard of care. Whether such a uniform standard would produce better medical care outcomes is unknown, and a uniform standard is intended neither by the legislation nor the PSRO program. In fact, the legislation requires periodic review of the norms and criteria on which the standards are based so that medical care will not become static.[43]

It is instructive to note that one of the principles evolved in malpractice law is the so-called "locality rule," which is based on the principle that a physician is to be measured by the standard of care of others similarly situated.[44] For a variety of reasons, this rule has been weakened, if not eliminated, in some jurisdictions.[45] With the advent of operational PSROs, the rule is likely to continue its demise as nationally generated data begins to form the basis on which decisions as to standards are made. Clearly, the structure of PSRO will affect malpractice not only through the determination of standards of quality care, but also by providing the potential for efficiently generated national data bases on which those standards may rest.

Incentives and Sanctions

A third primary difference between PSRO as a system of care review and other systems of review lies in the form of its incentives for compliance and its sanctions against noncompliance. Other than the right to continued reimbursement for services provided (which is more properly categorized under sanctions against noncompliance), the incentives to comply fall into the area of limitations on liability. These incentives to comply are of two general kinds: incentives to give information useful or necessary in the review process, and incentives to comply with the standards set by the PSRO. The limitations on liability for

Continued on page 257