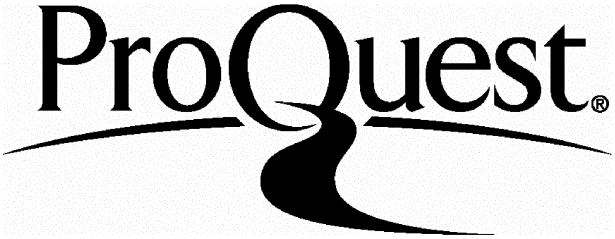
THE USE OF WALKING PARAMETERS IN QUANTIFYING DISEASE SEVERTY IN CERVICAL SPONDYLOTIC MYELOPATHY

Submitted for the degree of Doctor of Philosophy in the Faculty of Medicine (Clinical Studies)

By Anoushka Singh

Institute of Neurology
University College London
2002



Abstract

Currently there is much debate concerning the ideal management of cervical spondylotic myelopathy (CSM). There is uncertainty over which patients benefit from surgical decompression, the best timing for such surgery and the best type of surgical procedure to perform. These difficulties arise not only because of the unpredictable rate of disease progression and the variable results of surgical intervention but especially because of the inability to measure quantitatively, reliably and accurately the outcome of surgical intervention.

This thesis first uses a questionnaire to determine clinicians' current practice and attitudes in using quantitative measures of CSM severity and reveals their low level of use and a high level of disagreement over which are most appropriate. The demonstrated shortcomings lead in this thesis to the development of a simple walking test involving a timed walk over 30-m. A preliminary study of CSM patients studied before and after surgery reveals that such a walking test satisfies the criteria of a suitable scale of assessment and compares favourably with the most commonly applied existing rating scales. Actual use of the walking test is then demonstrated by a further study designed to look at the immediate and long term outcome of surgical decompression in CSM patients. Patients are followed up for 3 years after surgery and compared with a matched control group of CSM patients who did not undergo surgery. An immediate and maintained benefit is revealed compared to untreated patients who suffer an ongoing deterioration. These changes correlate with those recorded by a number of existing rating scales. Further analysis of the walking data shows that relatively *better* results can be achieved in older and more severely affected patients. Finally, walking data are compared with radiographic parameters and it is found that the presence of cervical cord T2-weighted MRI signal change correlates with a better outcome from surgery, but this correlation is not strong enough to make predictions in individual patients.

In summary, this thesis introduces and shows the usefulness of the walking test in assessment of CSM patients. The relative ease of use of the walking test may mean that it may find utility in normal clinical practice as well as in research trials.

Acknowledgement:

I am indebted to my supervisor Professor Alan Crockard for his advice, support, encouragement and generous sponsorship of my PhD. I am grateful to Mr Adrian Casey for his continued support during my research. My sincere thanks to Dr John Stevens and Dr Andrew Platts for their valuable assistance as co-authors.

I would gratefully like to acknowledge the Neurologists and Neurosurgeons at the National Hospital for Neurology and Neurosurgery, Queen Square, London who allowed me to study their patients. In addition I would like to acknowledge past and present Research Fellows for their moral support. Many thanks to the Surgical wards, Outpatients, Audio-visual, Radiology and Medical Records department for being patient with me. I would also like to thank Karen Cordell who helped me in numerous ways. Above all I would like to sincerely thank my patients without whom this study would not have been possible.

Declaration:

The work of this thesis was carried out in The Department of Surgical Neurology, at The National Hospital for Neurology and Neurosurgery, Queen Square, London. All the work was performed under the supervision of Professor Alan Crockard and in some cases with the collaboration of others as noted in acknowledgements. All experiments were approved by the local ethics committee and performed with the understanding and consent of each subject.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Definition

Degenerative disease of the cervical spine is termed cervical spondylosis. It begins in the intervertebral discs and results in changes to the surrounding structures. In the past, the condition has often been labelled as osteoarthritis, cervical spondylitis, herniated disc or chondroma, but the term spondylosis is generally used as it is a degenerative rather than a neoplastic or inflammatory condition.

1.2 Historical Perspective

The normal anatomy of the cervical spine, and in particular the intervertebral disc that is the focus of cervical spondylosis was first clarified by Vesalius over 400 years ago. However, the first pathological account linking clinical myelopathy with abnormalities of cervical anatomy was published much later when Key in 1838 (Key C.A., 1838) reported two patients with paraplegia in whom was found:

"A projection of the intervertebral substance or rather a posterior ligament of the spine, which was thickened (and ossified) and presented as a firm ridge which lessened the diameter of the canal by nearly a third".

The first operation for this condition, which is now termed cervical spondylotic myelopathy, was undertaken in 1892 by Victor Horsley. He performed a cervical laminectomy in a patient whose paraparesis had been precipitated by trauma and in this way decompressed a transverse ridge of bone that was found to be compressing the spinal cord at the level of the sixth cervical vertebra. Shortly following this, subsequent operations yielded pathological specimens of tissue that were repeatedly misidentified as

'chondromas' or 'neoplasms' of notocord origin. This was despite the fact that in the same year as Victor Horsley's first procedure, Gowers had already reported the pathological entity of vertebral exostoses; he described osteophytes protruding from the posterior surfaces of the vertebral bodies and encroaching upon the spinal canal to cause gradual cord compression. For many years, there remained little awareness of the importance of this process of degenerative *spondylotic* myelopathy. Attention focussed instead on "chondromas" or on "acute ruptured disc" as the primary pathological process (Stookey, 1928).

Russell Brain (Brain, 1948) finally set the emphasis again firmly upon cervical spondylosis, a process where disc degeneration, associated osteophytic outgrowths and changes in the surrounding joints and ligaments results in chronic spinal cord and root compression. He clearly distinguished this from acute, usually traumatic, rupture and protrusion of the cervical disc, a pathology more likely to compress the nerve roots than the spinal cord. Subsequently (Payne and Spillane, 1957), documented the importance of a congenitally smaller than normal spinal canal as an additional factor in the genesis of myelopathy in patients with cervical spondylosis.

These reports have been followed by a wealth of pathological, surgical and radiological articles on the subject, which have been summarised recently in a review by (Rowland, 1992).

1.3 Pathophysiology of Cervical Spondylosis

Cervical spondylosis constitutes a degenerative disease of the cervical spine whose major components include degeneration of the intervertebral discs and associated osteoarthritic changes of the intervertebral joints with bony overgrowths of the vertebrae

and ligaments in the form of osteophytes. The condition is now recognised to be extremely common in middle-aged and elderly populations (Lindsay, Bone et al., 1991).(Rengachary and Redford, 1991) In people above 50 years of age, around 40% have some clinical abnormality of the neck and using more sensitive radiological criteria, this figure rises to about 75% (Pallis C, 1954;Irvine, 1965;Hughes, 1965). By the age of 70, as many as 98% of people have evidence of degenerative changes in the cervical spine (Hunt, 1980;Montgomery and Brower, 1992). Its severity in the above described populations is of course highly variable and when mild, it may be asymptomatic or present simply with mild neck pain exacerbated by movement (Travell and Simons, 1983). Generally, such symptoms are treated conservatively with rest and analgesic drugs.

When the degenerative deformation is severe enough to compress the nerve root outlets from the spinal canal, it may result in a severe radicular pain radiating down the arm accompanied by weakness and sensory disturbance of the hand and arm. Since spondylosis is most severe at the lower cervical levels, the hand is characteristically involved with muscle wasting and sensory loss in the inner arm. Radicular symptoms may be treated conservatively in the same way as local pain or by anaesthetic injection around the nerve root as it leaves its exit foramen. Decompressive surgery for radicular symptoms is often inadvisable because it has an uncertain outcome and recurrence is common (Persson, 1997).

Finally, when the degeneration presses on the spinal cord itself within the spinal canal, spastic paraparesis below the neck may occur and over time may become complete and irreversible (Zeidman and Ducker, 1992). The process of cord compression and paraparesis is termed cervical spondylotic myelopathy (CSM) and it is the management of this stage of the disease that is the focus of this thesis.

The initial lesion in CSM, as in all forms of cervical spondylosis, is a tearing of the annulus fibrosis of the intervertebral disc, which results in a posterior bulging of the disc or actual extrusion or sequestration of disc nucleus material into the spinal canal. Both processes stimulate a fibrous and bony reaction, resulting in dural thickening and adherence to the posterior longitudinal ligament and in the formation of osteophytes and transverse bony ridges (fig 1.1). While lateral extrusion and osteophytic protrusion may result only in narrowing of the intervertebral spaces and spinal root compression (fig. 1.2b), midline disc material or osteophytic posterior longitudinal ligament will compress the anterior aspect of the spinal cord itself (fig. 1.2a). A congenitally smaller spinal canal makes certain patients much more susceptible to cord damage from a given severity of degenerative change (Payne and Spillane, 1957).

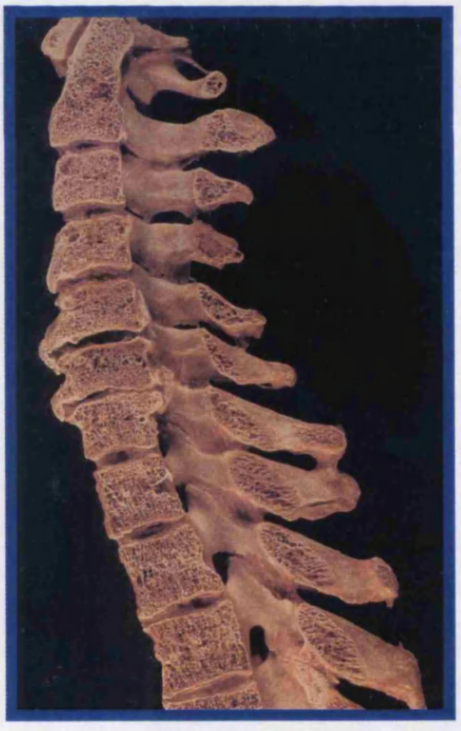


Figure 1.1 Photograph of a macerated specimen of the cervical spine in a 67 year old male. This photograph shows segmental narrowing of the spinal canal at levels of C5,C6 and C7.

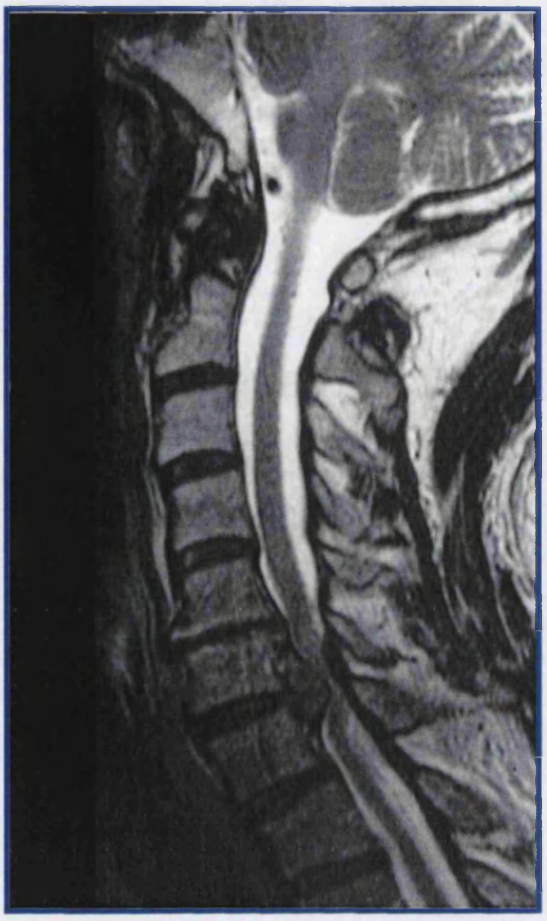


Figure 1.2a Sagittal MRI image showing large disc causing severe cord compression

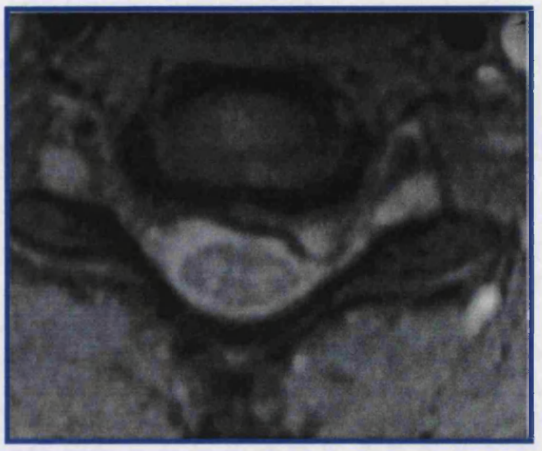


Figure 1.2b Axial MRI image showing lateral extrusion causing nerve root compression

Pathologically, the damaged cord shows areas of demyelination or focal necrosis in the posterior and lateral columns as well as loss of nerve cells in the grey matter (fig. 1.3). It is generally accepted that such changes result from direct mechanical compression between spondylotic bars anteriorly and the ligamenta flavum posteriorly (Stoltmann, 1964). However, some damage may also result from ischaemia, particularly from compromised venous drainage (Gooding, Wilson C et al., 1972;Gooding, Wilson et al., 1975;Hukuda and Wilson, 1972).

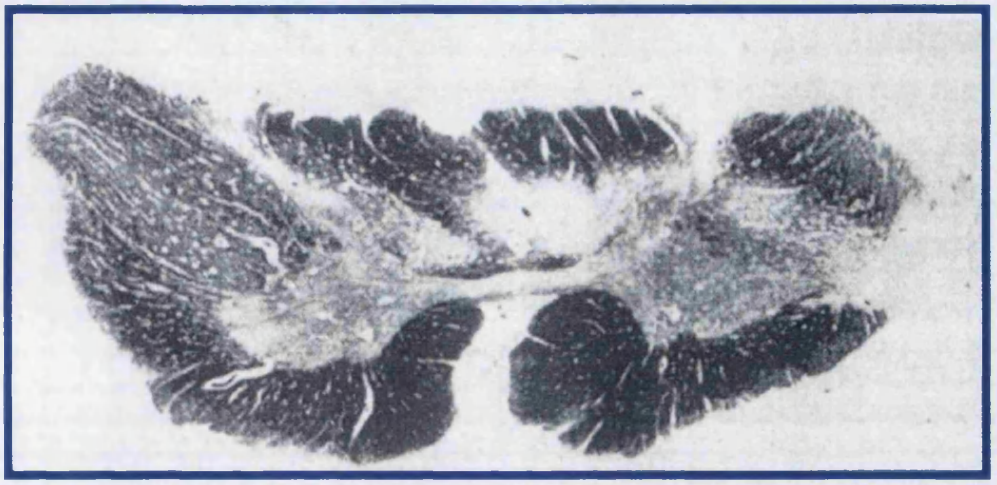


Figure 1.3 Section of the cord at the level of disc protrusion (C6/7) showing degeneration in the region of attachment of the ligamentum denticulatum on one side and degeneration in the ventral part of the dorsal columns chiefly on the same side (myelin stain).

The fact that the brunt of the damage is generally borne by the lateral aspects of the spinal cord (where the dentate ligaments attach to the cord) and, to a lesser extent, by deeper regions (Ogino, Tada et al., 1983) but not at the anterior and posterior margins, has led some authors to doubt the accepted view that damage results from direct compression (Breig, Turnbull et al., 1966). The lateral aspects of the spinal cord are attached to the dura by the dentate ligaments. The dura is relatively fixed against the bony walls of the spinal canal by the dural root sleeves held fixed within the nerve root exit foramina. Thus, if the spinal cord is displaced backwards by spondylotic bars growing from the backs of the vertebral bodies, this will draw the dentate ligaments taut, resulting in pulling against the lateral aspects of the spinal cord (Kahn, 1947). Detailed modelling of the various compressive and tensile forces (Levine, 1997) indeed reveals that the observed pattern of damage fits much better with the pattern of maximal shear stresses

produced by a tension mechanism than those that would be produced by a compressive mechanism. Such results are intuitively clear without modelling; the tension model predicts most damage at the lateral margins, where it is found to occur, while direct compression would result in superficial damage anteriorly and posteriorly. However, this tension hypothesis still remains contentious.

If the pattern of pathologically observed damage is mapped onto a diagram showing the anatomy of the long ascending and descending white matter tracts of the cervical cord (fig. 1.4), it is seen that the areas of increased damage correspond with the lateral corticospinal tracts (transmitting most of the signals for voluntary movements to the muscles) and the posterior and anterior spinocerebellar tracts (transmitting sensory information on e.g. posture and movement for processing by the cerebellum). These facts might prove useful in designing a study to test clinical and functional deficits in cervical spondylotic myelopathy. One might predict on this basis that measures of voluntary movement, balance and co-ordinated activity would be more sensitive than measures of, for example, bladder dysfunction, conscious joint position sense or temperature and pain sensation. Clearly, the exact anatomy of damage will be different in each individual patient, but the above pattern of severity may at least provide a rationale for selecting factors to monitor in an overall population of patients.

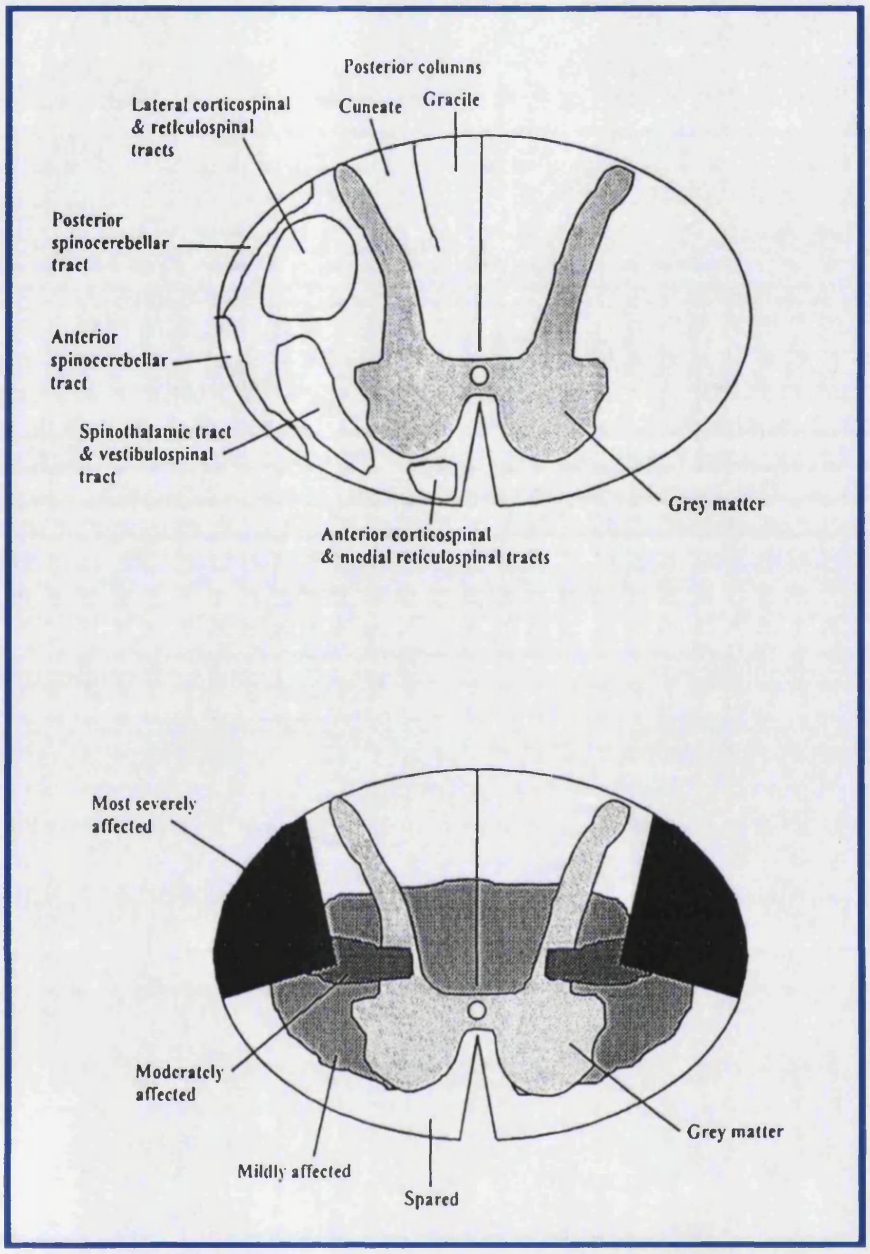


Figure 1.4 Areas affected in cervical spondylotic myelopathy.

In discussing the pathology of cervical spondylosis, it should be noted that other processes are important in resulting in degenerative cervical myelopathy in certain populations. Ossification of the posterior longitudinal ligament (OPLL), which runs lengthways along the vertebrae just anterior to the spinal cord, appears to be a hereditary

process relatively common in Asians, especially Japanese, and accounts for the bulk of cases of degenerative myelopathy in such groups (Ohtsuka, Terayama et al., 1986).

1.4 Clinical features of Cervical Spondylotic Myelopathy

The myelopathy associated with spondylosis of the cervical cord may present insidiously or with sudden exacerbations, perhaps related to neck trauma. The onset is typically in the 6th decade and males are more commonly affected. Although most patients suffer a relentless decline, some have long periods of stability. In fact, functional adaptation to fixed disabilities is sometimes interpreted as a period of improvement (Clarke and Robinson, 1956).

As mentioned above, the clinical features relate to the fact that most damage occurs in the lateral and posterior columns. When in a relatively early stage there may merely be slight stiffness of gait, often recognised as a specific problem rather late in elderly and generally infirm patients. As a result, early manifestations are perhaps more likely to include a mild *functional* gait deficit rather than abnormal neurological signs on examination on the couch. There may also be associated non-myelopathic spondylotic symptoms. Thus the classical triad of symptoms includes a painful stiff neck, brachalgia and spastic/ataxic leg problems. The first of these symptoms, a painful and stiff neck, is common and non-specific. Radiological findings on plain x-ray are even less specific (Pallis C, 1954), in a survey of 50 patients over the age of 50 and without neurological complaint, found that 75% showed radiological narrowing of the cervical spinal canal due to posterior osteophytosis or narrowing of the intervertebral foramina due to osteoarthropathy at the neurocentral and apophyseal joint. About half the patients with radiological abnormalities showed physical signs of root or cord involvement, even

though they had no symptoms other than neck ache.

In addition to gait problems, other more specific myelopathic symptoms include impaired lower limb (particularly dorsal column) sensation, paraesthesiae and Lhermitte's phenomenon. Involvement of sphincters, based on the patient's perception, is unusual at presentation (Brain, Northfield et al., 1952;Arnold, 1955;Clark, 1988). Typical *signs* of cervical myelopathy include pyramidal weakness, sensory loss below the neck and a broad-based, and spastic gait (Carr, 1992;Montgomery and Brower, 1992).

Diagnosis is confirmed nowadays by magnetic resonance imaging; myelography combined with CT scanning is less used.

1.5 Treatment of Cervical spondylotic myelopathy

The accepted mainstay of treatment for spinal cord compression resulting from cervical canal stenosis is surgical correction of the deranged bony anatomy. The first operation for spinal cord compression was probably performed by Victor Horsley in 1892 (Horseley, 1892). Since that time, surgeons have used an increasing variety of operative techniques to decompress and stabilise the cervical spine, the ideal method obviously being the one that provides most direct and safe removal of compressive pathology and which thereby prevents progression of disabling neurological deficit, alleviates pain and helps a return towards normal functioning.

Initially, decompression was performed by a posterior surgical approach involving a laminectomy. Posterior approaches, although less frequently performed in contemporary times, have nevertheless remained of value for treatment of myelopathy and radiculopathy due to cervical spondylosis (Epstein, Carras et al., 1969;Fager, 1973;Piepgras, 1977). In laminectomy the aim is to remove those posterior elements

which cause compression of the dura and its contents, including the hypertrophied and infolded ligamentum flavum. The advantage of the laminectomy procedure is that if a comorbid radiculopathy is present, a foraminotomy at an appropriate level or levels can be performed at the same time. The disadvantages are the risk of causing instability, although unlikely if the facet joints are preserved, and the failure of the technique to enable easy access to anterior structures.

An alternative to standard laminectomy is a cervical laminoplasty, where the vertebral canal is enlarged without complete removal of the spines and laminae (Hirabayashi, Watanabe et al., 1983). This should better preserve supportive function of the vertebral column posteriorly (Fearnside, 2000). There is no general agreement on one posterior procedure having a clear outcome advantage over the other, mainly because there has been no long- term follow-up controlled comparison.

During the past quarter of a century, the anterior approach has become popular, the posterior approach being reserved largely to treat multi level cervical spondylosis, and unilateral, far lateral, and foraminal disc herniations. This is perhaps because it is considered that the disc anterior to the cord is the major pathology (Brain, Northfield, and Wilkinson, 1952;Hashizume, Iijima et al., 1984;Gooding, Wilson C, and Hoff J, 1972). The three most common anterior operative techniques were described by Clowards, Robinson and Smith, and Bailiey and Bagley (Clowards, 1958;Robinson and Smith, 1955;Bailey and Bagley, 1960). All of them allow direct access to vertebral bodies and intervertebral discs. In general the perceived benefits of an anterior cervical decompression (ACD) with fusion are prevention of further osteophyte formation, regression of existing osteophytic spurs, disc space distraction which reduces ligamentum flavum buckling and enlargement of the neural foramina with consequent nerve root

decompression. Further advantages of this approach are the ease of positioning of the patient and minimal muscle trauma. The major disadvantage is the potential for injury to the soft tissue structures of the neck, including the carotid artery, recurrent laryngeal nerve, the trachea and the oesophagus.

Early complications of either surgical approach include infection, haemorrhage and immediate worsening of neurological state, said to be more frequent after posterior compared to anterior procedures. Late complications (after the 1st week) are neurological deterioration, wound pain, cervical instability and kyphosis.

None of the available techniques have been rigorously compared in clinical trials with each other (Yonenobu, Fuji et al., 1985) or with conservative management, and the best course or choice of treatment strategies varies with the bias of the clinician. Many separate published reports have described the results of individual techniques, but authors have varied greatly in their selection of patients included in such studies, depending upon the definition of cervical spondylosis (Payne and Spillane, 1957;Campbell and Phillips, 1960;Dunsker, 1981). Studies may include acute herniated discs with chronic spondylotic bars (Phillips, 1973) or combine myelopathy and radiculopathy patients together (Clarke and Robinson, 1956). Thus direct comparisons between studies is difficult. Moreover, some studies use only radiological criteria to make the diagnosis and miss out reference to clinical significance (Freidenberg and Miller, 1963;Hitselberger and Witten, 1968;Teresi, Lufkin et al., 1987;Lawrence, Bremner et al., 1966). Conversely, other patient series lack an explicit description of imaging findings.

It may be that, with increased information about the results of various procedures, it will be seen that different *anatomical* problems (on imaging) lend themselves to different operative procedures. In addition, different *clinical* presentations might result in

different benefits from different operations, since different types of damage are likely to affect different anatomical locations of the cord and discrete spinal neural tracts will have different effects on neurological function.

1.6 Selection of Patients for Surgery

An important point to consider is that not all CSM patients undergo immediate surgery. As discussed above, the insidious progression of the condition means that there will be a continuous variation of clinical severity from no disability to irreversible paralysis. There is also a similar variation in pathological or "operative" severity that, at least in anecdotal surgical practice, often corresponds rather poorly with clinical severity. Thus many patients do not and should not undergo surgery.

Clearly, therefore, some kind of *selection* process must take place in determining which CSM patients will benefit from an operation as well as in determining which operation should be performed. In general, clinicians rely on specific symptoms, such as difficulty with gait or urinary difficulties, together with specific findings on clinical examination and radiological imaging, not only to make the diagnosis but also to identify the most severe forms of cervical spondylosis and to decide when surgery is appropriate (Hadley and Sonntag, 1996). However, different clinicians appear to vary greatly in their selection practices for spondylosis surgery and it is possible that a considerable number of patients are unnecessarily operated upon, while others are operated upon too late or not at all. The former error may lead to severe morbidity because of the risk of complications during the anaesthetic or operation and because of the problems of rehabilitation from major surgery. The latter error may leave a patient permanently and potentially avoidably severely disabled.

As a result of these difficulties in selecting patients for surgery, clinicians are turning increasingly to quantitative or semi-quantitative guidelines such as ratings of neurological impairment (Ranawat, 1979) or functional disability (Steinbroker, et al., 1949). In other words, the selection for operation, rather than being based solely on personal clinical judgement, should now be performed more scientifically. However, to date, the same problems that have dogged a proper comparison of operative techniques have also dogged a determination of the best timing for surgery. In addition, the means of assessing clinical status, both pre-treatment and in the outcome phase have not been standardised across trials (Rowland, 1992;Symon and Lavender, 1967;Zeidman and Ducker, 1992).

1.7 Outcome Measures

The identification of a need to find a common objective means to assess the effect of surgery is not an isolated break in strategy, but reflects the general trend of modern science and medicine. To rationalise such management philosophies, one might say that the traditional approach has been based on Deductivism. Hume, an 18th century Scot considered the father of British Empiricists, described the basis of this philosophy thus:

"Propositions...are discoverable by the mere operation of thought, without dependence on what is anywhere existent in the universe."

(Hume 1751 cited by Tomassi) (Tomassi, 1995).

However, Hume felt that the modern scientific method had superseded such ideas:

"Men are now cured of their passion for hypotheses and systems in natural philosophy and will hearken to no arguments but those which are derived

from experience."

Unfortunately, three centuries later, much of modern medicine is still entrenched in Deductivist reasoning. Taking the case of management of CSM, there remains a strong tendency for the subconscious assertion: "In cervical spondylosis, there is compression of the spinal cord – therefore operative decompression is beneficial in such patients". In the modern era, we should be justifying our management of disease by multiple observations and measurements.

Recent NHS policy trends should accelerate this change to an empirical approach of management. Government agendas focus on the demonstration of provision of first class care, cost effectiveness, continuous revalidation of clinicians, Clinical Governance and a search for objective means of assessment of clinical outcome.

Implementation of this approach to CSM management should involve a formal and scientific study investigating a large number of patients and categorising the patients' condition in a reproducible and simple manner before and after surgery. The results of performing surgery at different levels of progression of the condition could be analysed so that an optimal timing is determined. Provided there are no ethical constraints, this could also be assessed by comparing patients undergoing such surgery with others who for one reason or another did not have surgery. Obviously, conducting trials on a blinded basis is not appropriate for these studies!

The assessment of patients constitutes a measurement of severity and of outcome. This measurement should be standardised. A standardised measure of outcome is critical since it is this that allows quantification and reproducibility. This standardisation often takes the form of *categorisation* of severity, so that patients of similar severity category can be identified and compared. Thus the results of a research study can be transferred to

a normal clinical setting so that clinicians can draw on the experiences of the study and apply measurement criteria and severity categories both to individual patients and as part of audit of overall practice.

1.8 Ratings scales

The measurement of health related outcomes such as disability and quality of life are essential factors in the evaluation of therapeutic efficacy. Categorisation is normally done in the form of a rating scale (Nurick, 1972;Hirabayashi, Watanabe, Wakano, Suzuki, Satomi, and Ishii, 1983). This can relate to the severity of radiological imaging, the severity of the neurological examination findings or to the severity of dysfunction of the patients' daily activities. When one is devising a scale for quantifying or categorising a condition's severity, there are many factors one must take into account. An ideal scale should:

1) Be quantifiable. An **ordinal scale** merely lists a finite number of non-numeric items in order of severity and assigns an arbitrary number (MacKenzie and Charlson, 1986) while an **interval scale** is a proper continuous range of severity (Martin, Leltzer et al., 1988). A **ratio scale** is an actual measurable number (Froberg and Kane, 1989;La Rocca, 1989). The latter two are more quantifiable in that the number has more meaning and is more amenable to statistical comparison.
2) Have a suitable distribution. The range of patients' values (scores) should ideally be spread evenly or normally throughout the range of a scale. If a scale ranged from 0 – 100 and nearly all patients fell in the range of 95 – 100, the scale would have little useful meaning.
3) Be valid in that it measures what it purports to measure (Wassertheil-Smoller,

1995).

4) Be responsive to change. If one wishes to measure the change in the scale value over time or after an operation, the scale should be responsive to change, (MacKenzie and Charlson, 1986;Guyatt, Walter et al., 1987) preferably in a graded manner.

5) Be easy to perform, preferably requiring no special training and a minimum of time.

6) Have good intra and inter-observer reliability, yielding the same results for repeated uses under the same condition. A reliable measure produces results that are consistent, stable over time, and reproducible. Inter-observer reliability is the agreement between two or more raters and may be assessed by examining the correlation between ratings obtained from independent observers. Intra observer reliability, which is the agreement between two ratings made by a single observer on the same patient, is assessed using the same statistical methods.

7) Have good internal consistency. This is a measure of homogeneity of the scale. If a multi-part questionnaire included a number of questions all designed to measure the same factor, eg mobility, then the different questions should yield consistent answers. This may be assessed by the statistical Cronbach's alpha score (Cronbach and Meehl, 1955).

8) Be one-dimensional. The scale should reflect severity due to cervical spondylosis and not be dependent on extraneous or confounding factors or represent the combination of many interacting factors that do not range from mild to severe in a simple manner.

9) Be relevant. The scale should reflect severity in terms of what is actually

important for the patient rather than for example simply something which gives good statistical results for surgery.

A number of studies have developed and explored the use of such ratings scales (Table 1.1), and some are indirectly or directly applicable to CSM. Such scales are generally divided into those which score neurological impairment e.g. weakness, sensory impairment and spasticity, those which score disability e.g. limitation of mobility, inability to feed and dress oneself, and those which score handicap, the way a patient's functioning as part of society is affected.

1.8.1 Impairment Scales

Impairment scales for conditions related to cervical spondylosis include the American Spinal Injuries Association (ASIA) scale for spinal cord injury, which is based on the international standards for Neurological and Functional Classification. This has five categories of progressively decreasing severity and spinal level of motor and sensory dysfunction (appendix 1). However, the ASIA score has been criticised on a number of counts for being too detailed, time consuming and monotonous. The most important problem for application to CSM is that it is aimed at traumatic spinal injury patients. A scale for a slowly progressive condition such as CSM must have different qualities, such as sensitivity to change, to assess the correct time for operative intervention; the purpose is therefore different from that for assessment of severity of acute injury. Moreover, a problem that may be applied to many scales is that they have simply not been validated for certain groups of patients. A scale validated for one condition, such as spinal trauma, cannot be assumed to be valid for another, such as CSM. In fact, some authors suggest that any studies based upon any assessment tools that have not had their reliability and validity tested for that group of patients should simply be discarded (Wade, 1992).

Ranawat's neurological classification for cervical cord compression includes five categories of impairment (appendix 2). These are based on subjective weakness, hypereflexia, sensory changes, 'long tract signs' and the ability to walk. However, some of these factors are clearly mixing functional as well as impairment aspects.

A problem with scales such as the Ranawat scale is that they are poorly quantitative (being ordinal rather than interval) with very few and largely arbitrary categories. The sensitivity to change is likely to be poor since one category covers a huge range of actual severity. Since the information provided from a detailed clinical examination is so much greater, the scales are likely to be more applicable to statistical grouping than the management of individual patients. For example, Ranawat's scale distinguishes subjective weakness and tingling from objective weakness and clear signs of cord compression; such a distinction is so obvious in an individual patient as to be of limited use.

Category	Scale	Comments	Reference
Japanese orthopaedic association score for cervical myelopathy	Motor/sensory/bladder functions assessed (normal 17) Recovery rate (%)	No formal tests of reliability or validity	Hirabayashi et al. (1983)
European myelopathy score	Scores range from 5 (worst) to 18 (normal)	A scale adapted from the JOA for Western use that also includes pain assessment	Herdman et al. (1994)
Myelopathy disability index	11 questions; graded response (0-3). Disability expressed as a percentage.	Derived from Stanford HAQ. Reliability, validity and responsiveness documented for rheumatoid disease*. Repetition on limited aspects of function. Also validated for cervical spondylotic myelopathy†	*Casey et al (1996) †Singh and Crockard (2001)
Nurick grades	Grades ranging from 1 (normal) to 5 (worst)	Assesses walking only - ignores hand function. Empirical selection of grades. No validation	Nurick (1972)
Ranawat neurological classification	I (normal) II (subjective myelopathy) IIIA (objective weakness) IIIB (unable to walk)	Simplistic, poor discrimination between II and IIIA	Ranawat et al (1979)
Pain: Visual analogue scale	0-10 ruler	Simple and reliable	Huskisson (1974)

Table 1.1. Scales for assessment of neurological disease in the cervical spine

1.8.2 Functional Scales

Functional scales are perhaps somewhat more helpful in quantifying and rationalising factors important to the patient that might otherwise receive less attention. However those functional scales so far used suffer to a greater or lesser extent from a high degree of subjectivity, leading to potential inaccuracy and poor reliability. Nurick (Nurick, 1972) has developed a very simple functional scale that concentrates on lower limb mobility (apart from root pain for mildest involvement) (see appendix 3). Since then, the Japanese Orthopaedic Association (JOA) scale (Hirabayashi, Watanabe, Wakano, Suzuki, Satomi, and Ishii, 1983;, 1994)has been developed which is more detailed and concentrates on different aspects of function. It mixes impairment with disability and splits into four categories (For details please see appendix 4.):

1. Motor dysfunction of arms
2. Motor dysfunction of legs
3. Sensory deficit
4. Sphincter dysfunction

Each of these categories is graded 0 (most severe) to 2, 3, or 4 (moderate, mild, normal) and the total score is added. The score has been felt to be unsuitable for use in various centres in other parts of the world as unfortunately it was designed specifically for the needs of the Japanese population, where patients' hand function was assessed by asking if they could use chopsticks. Moreover, the scale was partly used for assessing clinical progression of disease related to ossification of the posterior longitudinal ligament of the cervical spine, a condition that is very common in Japan but not often a major feature of spondylosis elsewhere in the world. Thus the scale may suffer both from lack of relevance and lack of validity.

The European Myelopathy score (Herdman, Linzbach et al., 1994)(Appendix 5)

evolved as a result of perceived problems with the JOA scale and was designed to be easy to use. It is purely functional in its assessment, meaning that the criteria explored allow a critical evaluation of cervical myelopathy in terms of the actual severity of symptoms suffered by the patient in their everyday activities.

Odom's criteria (Odom, Finney et al., 1958)are very simple and roughly functional measures that specifically deal with the results of surgical intervention (Appendix 6). However, the categories are so general as to simply assign patients into good, bad and intermediate groups.

Rather than designing a functional scale specifically from scratch, an alternative strategy in applying a scale to CSM is to use a general scale that is already well established and has at least been validated for general use. If there are found to be limitations for specific use in CSM, then the scale may at least provide a starting point for specific modifications. One of the best known neurological functional scales is the Barthel Activities of Daily Living index (Appendix 7). This scale can be applied to any neurological condition but is quite complex to conduct. A variation of this scale is the Health Assessment Questionnaire (Appendix 8) which takes twenty questions taken from the Barthel scale and patients can conveniently answer the questions themselves by questionnaire.

In an effort to develop a scale for myelopathy (Casey, Bland et al., 1996),applied this Health Assessment Questionnaire scale to rheumatoid arthritis patients. They then isolated what they felt were the 10 (or 11) most important questions specific to cord compression and created a new questionnaire called the Myelopathy Disability Index (MDI) (see Appendix 9). They found that both the Health Assessment Questionnaire and their modified MDI had a good correlation with Ranawat's impairment scale. The MDI was also sensitive enough to show that patients operated upon when still quite good

according to their score did significantly better post-operatively, as judged over time by mortality rate, by the similar HAQ disability questionnaire and also as judged by the Ranawat impairment scale. They performed statistical analysis to show that their measure satisfied many of the criteria of a good scale, namely a good wide distribution, sensitivity and responsiveness to change and one dimensionality. It was also easy to perform and, since it was a functional measure, likely to be relevant to the patients' well being. However, to be applied to myelopathy resulting from cervical spondylosis rather than rheumatoid arthritis, a separate analysis and validation would have to be performed.

1.8.3 Handicap scales

No handicap scales have yet been developed for CSM, but some general health questionnaires have been validated for use in certain other neurological conditions. The Short Form 36 Health Survey (SF36) (Ware and Sherbourne, 1992)(appendix 10) is a very detailed mixed handicap and functional disability score, consisting of a patient-filled questionnaire and a complex scoring system. Economists and the Government, in determining cost-benefit ratios of various medical interventions, use the Quality Adjusted Life Year Score (QALY) (Bush, Anderson et al., 1982) for allocating healthcare resources. One QALY is a year of full life quality and so a degree of handicap is scored on a continuous scale as a fraction thereof. This is the first measure discussed so far that has proper qualities of an interval rather than ordinal scale, allowing more quantitative analysis to be performed. Thus the QALY is designed so that it can be multiplied over years so that one can make an attempt to compare a good quality of life over a short time with a poorer quality of life over a longer time.

1.9 Aims of the current studies

Most of those who have worked in clinical science and in health care in general have found that clinicians are strongly swayed by their own clinical experience of

severity of neurological impairment. Thus a criticism often levelled at functional scales is that they lack the objectivity of a clinical neurological examination. For example, a scale based on patients' reporting of their own symptoms might be coloured by their general psyche and by how urgently they perceive they need an operation. It would be advantageous therefore to achieve further rationalisation of severity beyond the above scales of functional impairment. Thus there is a real need for objective measures of severity that are functionally relevant, provided they represent or at least mirror the full picture of disease severity.

The primary aim of this thesis is to address such needs and thereby improve on existing practice for surgical treatment of CSM. The current practices of clinicians in their determination of surgical outcomes are first surveyed, together with their opinions regarding the criteria fulfilled by an ideal scale. There is little point in developing a complex scale which undergoes extensive validation, only to find that it at best is only ever used as a research tool and is never actually adopted in clinical practice, or at worst is simply one of many different research tools each used by competing research groups.

Consideration has been therefore given to adopt a good functional "test" of overall spinal cord integrity. From anatomical considerations of the spinal tracts most severely involved pathologically, it seems that measures of voluntary movement, balance and co-ordinated activity would be better measures than bladder function, conscious joint position sense or temperature and pain sensation. A measured walking test was therefore selected as a prime candidate for such a gauge of CSM severity.

Walking is a complex activity requiring voluntary control as well as processing of co-ordinated muscle activity. A walking test involving a timed walk would be directly quantifiable and would therefore qualify as a ratio scale. It is also very functionally relevant for the patient, and for clinicians who already in fact use a qualitative assessment

of walking as a "rough and ready" measure of spastic paraparesis resulting from cord compression. A walking test, which constitutes an important functional measure and yet is also closely and directly related to the neurological impairment of spastic quadraparesis, may thus have the advantage of serving to bridge the gap between disability and impairment measures. It is possible therefore that such a test may provide a very simple to perform and all-embracing measure of severity that surgeons feel confident in using, since it relates to their clinical training in performing neurological examinations and assessing impairment. It is hoped that, as a result, the test may receive widespread acceptance.

Short 10-m walking tests have already been used to a certain extent in functional assessment of other neurological conditions (Holden, Gill et al., 1984;Wade, Wood et al., 1987). In attempt to provide a more accurate measure that includes the functionally relevant component of fatigue, this thesis has selected a 30-m timed walk with one turn as the primary walking measure.

Much of the work of the thesis involves assessment of the responsiveness, validity, reproducibility and practical usefulness of this walking test according to the criteria described above for scales in general. The tests are conducted upon patients pre and post operatively at different stages of progression of cervical spondylosis to attempt to determine which patients will derive most benefit from surgical cervical cord decompression and also the optimal timing for such surgery. These results are compared both with normal controls and with unoperated CSM controls to make the vital comparison of intervention with the natural history of the condition.

In addition, the same selection criteria of ease of use, quantifiability and functional relevance are applied to other secondary measures, such as pain as measured on a visual analogue score (Huskisson, 1974) and certain easily measurable features on

magnetic resonance imaging of the cervical cord.

Finally, it is anticipated that the walking and other tests, as well as being used independently, might be weighted and amalgamated to provide a comprehensive measure of severity. Such a system could provide the basis for research comparing various treatment strategies for CSM as well as being applicable to the standard implementation of these strategies in normal practice.

2. METHODS

This chapter describes methods common to the bulk of the thesis. Details specific to individual studies are described with those studies.

2.1 Patient Selection

Patients used for the study were in general those consecutively referred to the neurosurgeons at the National Hospital, Queen Square, London, UK (NHNN). The patient mix reflected the referral pattern to this hospital; this mix was heavily biased toward patients with chronic rather than acute disease. The rare patient with acute myelopathic deterioration from cervical spondylotic myelopathy was sometimes surgically decompressed as a relative emergency without referral for pre-operative assessment. It was considered entirely appropriate that these studies focussed upon chronic cases since this is the most common presentation and the one where decisions regarding whether or not to operate and the timing of such surgery are more difficult.

The patients were operated upon by six different neurosurgeons. Those involved with walking and measurement scale assessment did not influence operative decisions or the surgical approach. These studies were therefore prospective and observational.

2.2 Ethics

Ethical Committee approval was obtained for the study and informed written consent was obtained from each patient under the guidelines of the local Hospital Policy and Procedures. All patients were supplied with an information sheet describing the study. In one study, a control group of myelopathic patients who did not undergo surgery was assessed and followed up. These patients had all been referred with CSM. The decision not to operate was made by the patient in consultation with his own physician or surgeon, not one that was made by anyone directly involved in the study. However, they were all patients that some surgeons would have been prepared to operate upon. Typically, the non-operated patients constituted those who wanted to avoid surgery and

its attendant risks and felt that they could get by with their current level of difficulty. They may also have been advised against surgery by their physicians. The patients nevertheless agreed to have regular monitoring of their walking parameters. All such patients were explicitly informed that they could change their mind about surgery at any stage. In fact a third of these patients did later change their mind and elect for surgery.

2.3 Diagnosis of CSM

All patients had the diagnosis made by experienced neurologists or neurosurgeons and corroborated by magnetic resonance imaging. None had previously undergone neck surgery or had any other pathology that might have resulted in functional impairment.

2.4 Procedures

Patients were assessed on a number of measures pre-operatively and at defined times post-operatively. Pre-operative assessments were carried out a few days before surgery.

2.4.1 Walking Test

Patients and controls walked on a smooth flat surface over a ward corridor for a measured 30-m distance (15-m there and back with one turn) (fig. 2.1). The time taken was recorded using a stopwatch and the number of steps taken was counted (Singh and Crockard, 1999). The patient was requested to walk at his maximum comfortable speed (with any normally used walking aids). They did not run. For initial reliability checks, each trial was performed three times. There was at least a ½ hour rest between trials.

Gait speed is dependent on a variety of factors such as cadence, stride length, balance and endurance. These are all factors that are affected by damage to the long pyramidal and sensory tracts so that this functional and quantitative measure is likely to reflect impairment and underlying pathologies. The fact that patients were allowed to use aids is making it behave more like a functional measure.

Fig 2.1 A patient taking part in a 30-m walking test at the postoperative assessment phase. The floor area is marked for 15-m there and back which includes one turn. The stopwatch used for the assessment is displayed at the top right hand corner of the picture. The number of steps are also recorded.

2.4.2 Pre- and Post-operative Scale Scores

The scores on a number of scales were recorded at the same times. They were performed by the patients under supervision. (Many of the scales have been validated for completion by both assessors and subjects.) The exact details of the scales are listed in the appendices 2, 3, 4, 5, 6, 9, 10 and 12).

2.4.3 Pain Analogue Score

In one study, a pain analogue score (Huskisson, 1974) was completed by the patients. Each patient was given a horizontal unmarked line where the left end represented no pain and the right extreme the worst pain imaginable. They made a mark on the line that

corresponded to their own severity of pain over the last few days. This was completed separately for neck pain and for arm pain.

2.5 Analysis of Scales

Statistical analysis was performed using Microsoft Excel and the Statistics Program for Social Sciences (SPSS) package version 10. The specific tests used are described in the relevant sections. In assessment of the qualities of different scales, various criteria were used which are defined below.

Sensitivity to change

It is clearly desirable for a scale to show a large sensitivity to change. Generally, the significance of changes over time or resulting from surgical decompression were analysed by repeated measures analysis of variance, paired samples t-tests or by the corresponding non-parametric tests. However, this does not reveal the *magnitude* of the change. This was quantified by calculating the normalised change, the mean of the differences following surgery for the subjects divided by the overall median of the pre- and post-operative scores (mean of (pre-op score minus post-op score)/ median of all pre- and post-op scores). (The mean of differences could be used because these differences followed an approximately normal distribution, even though the scores themselves may not have.)

Absolute Sensitivity

It may be desirable to have a high sensitivity to distinguish different absolute levels of severity between patients in the sample group, as well as sensitivity to changes following surgery. Absolute sensitivity was quantified by the coefficient of variation (the interquartile range divided by the median or the standard deviation divided by the mean).

Internal Consistency

It is important to attempt to determine the internal consistency or reliability of a scale. For a multi-part questionnaire, this is like a measure of reliability of response

across the different questions within a questionnaire. If different questions are attempting to measure the same parameter e.g. walking, then there should be consistent scoring within patients. This is measured by Cronbach's alpha, a normalised measure of correlations between multiple components of a scale. A score of 1 indicates a perfect correlation and zero no correlation.

Reliability

This is a measure of repeatability of a scale. A scale should be able to be measured on different occasions (and by different assessors where relevant) and give the same values provided the patient's condition has remained unchanged. This has already been validated for the previously used scales. For walking data, between-trials analysis of variance was used. For radiographic data analysed on two occasions by two assessors, Cronbach alpha was used in the same way as used above for different scale components. A score of 1 indicates a perfect reliability match.

Validity

To explore the validity of the different scales, they are generally compared with a "Gold Standard" test. In the absence of such a standard, scales are generally compared with each other using correlation coefficients. Kendall's rank coefficients were used for the non-parametric data.

3. Clinicians' Current Methods of Assessment of Myelopathy Severity

3.1 Summary

There is considerable uncertainty concerning patient selection and operation timing for cervical spondylotic myelopathy (CSM). Attempts have therefore been made to quantify CSM severity to provide an adjunct to clinical decision-making. The aim of the study described in this chapter was to determine, by means of a 7-item questionnaire, the attitudes of clinicians regarding the importance of quantitative assessment in normal management of CSM, their actual use of such assessment and how current scales fall short of ideal.

Quantitative assessment was regarded by 117 respondents as being almost as important as history, imaging and clinical examination; overall function and rate of progression were also highlighted as major factors. However, only 4 believed there was a "gold-standard" scale and the actual use of scales was low and fragmented. There were clear differences between specialities in both attitudes and practice.

The discrepancy between the importance attached to quantitative measurement and its actual use suggests that current scales are under-utilised or unsuitable for clinical practice. A new easy-to use scale may be required that better reflects clinicians' attitudes.

3.2 Introduction

A major difficulty with the management of cervical spondylotic myelopathy (CSM) is that its often slow and insidious course coupled with the potential hazards of surgery can result in uncertainty concerning the optimal operative procedure to perform and the optimal stage in the disease at which to perform the procedure. Current assessment of CSM generally involves subjective and non-standardised assessment of patients (Lunsford, Bissonette et al., 1980a;Harsh, Sypert et al., 1987). Partly as a result, clinicians vary greatly in their selection practices for surgery (Hadley and Sonntag, 1996;Sypert, 1992;Allen, 1952;Bakay, Cares et al., 1970;Gonzales-Feria, 1975;Sypert and Arpin-Sypert, 1996;Alexander, 1996).

The aim of this study was therefore to survey, by means of a questionnaire, clinicians' current attitudes to assessment of CSM severity. It was considered important to look at practices of all doctors potentially involved, including general practitioners and geriatricians, rather than just operators, because such clinicians determine initial referrals. In addition, the actual current use of quantitative assessment was surveyed. Finally, if there was a lack of use of quantitative measures, an attempt was made to ascertain in what way the existing measures do not meet clinicians' requirements.

Conducting such a survey before embarking on an original project looking into quantitative assessment of CSM would hopefully provide some initial insights into actual use of scales and into the properties of a scale that would make it most useful to clinicians.

3.3 Experimental Protocol

Two hundred postal questionnaires (see Appendix 11) were sent out in July 1998 to randomly selected UK clinicians involved in management of CSM. Replies included Neurologists 15%, Neurosurgeons 31%, Orthopaedic-spinal surgeons 26%, Rheumatologists 7%, Care of the Elderly Physicians 6% and General Practitioners 15%. This distribution is shown in fig 3.1.

The main points of interest were: (i) the relative importance of different methods of CSM severity assessment, (ii) the important qualities of a scale, iii) whether or not they considered a 'gold standard' scale to exist already, (iv) their actual use of scales in academic or clinical practice and (v) other criteria they actually used for assessment.

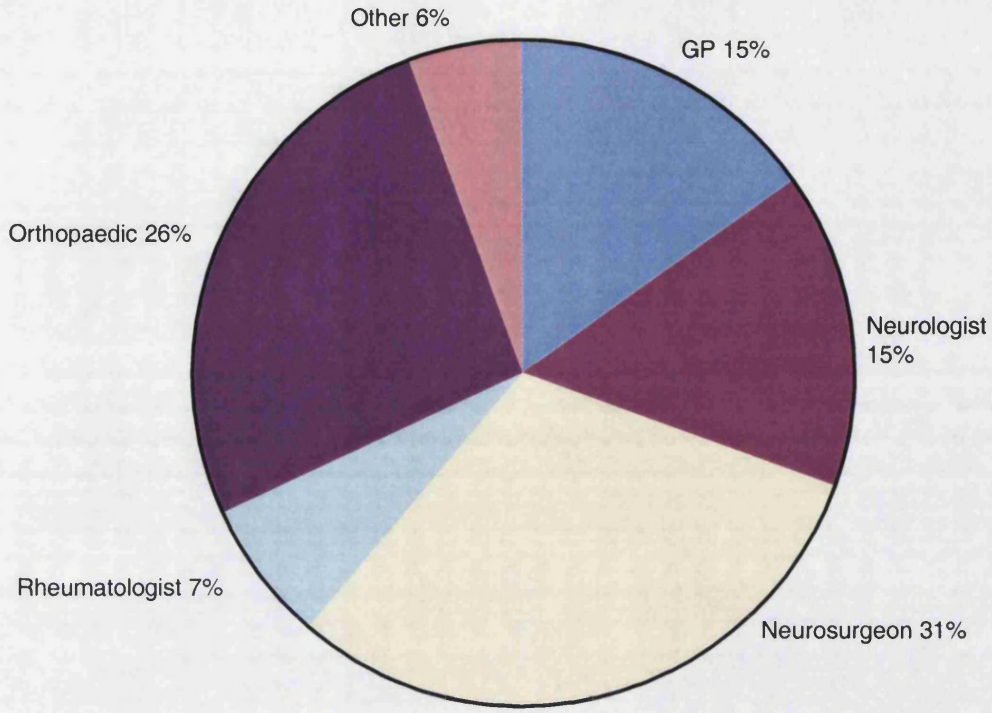


Figure 3.1 Pie chart showing the clinicians that responded to the questionnaire.

3.4 Results

Responses to the questionnaire were received from 65% of clinicians (n=117).

3.4.1 Closed questioning most important criteria for assessment. (Question 2).

History and imaging were jointly the most frequently indicated single most important criterion for determining CSM functional severity (fig. 3.2).

In addition to considering the most single important criterion, the overall *index of importance* of the five parameters (history, examination, imaging, quantitative measurement or "unknown/ don't know"), was determined by requesting the clinicians to rank the parameters in order of importance, 4 being the most important and 0 the least. The scores of all the clinicians were then summed for each parameter and divided by the number of clinicians to give a normalised measure where, for example, a final result of 4

would indicate it had been ranked top by each clinician. Data analysed in this way revealed that quantitative functional assessment and clinical examination were ranked almost as important as history and imaging (fig. 3.3). In other words, while not as often ranked top, these measures were often ranked 2nd or 3rd rather than last. Thus, overall, no parameter stood out as being very important or very unimportant.

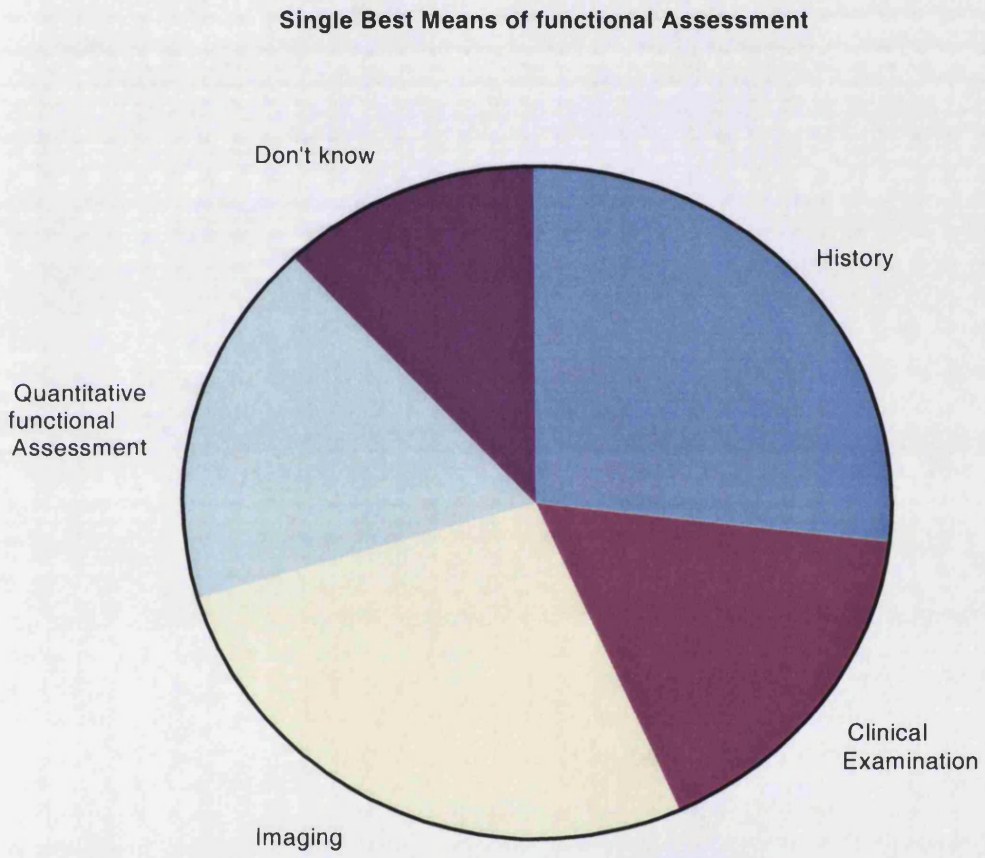


Figure 3.2 Clinical History and Imaging were both considered the single most important criteria for assessing functional severity.

Further analysis revealed that inter-speciality differences existed in attitudes regarding the relative importance of the different factors (question 1). In particular, GPs were unsure about CSM assessment (fig. 3.4).

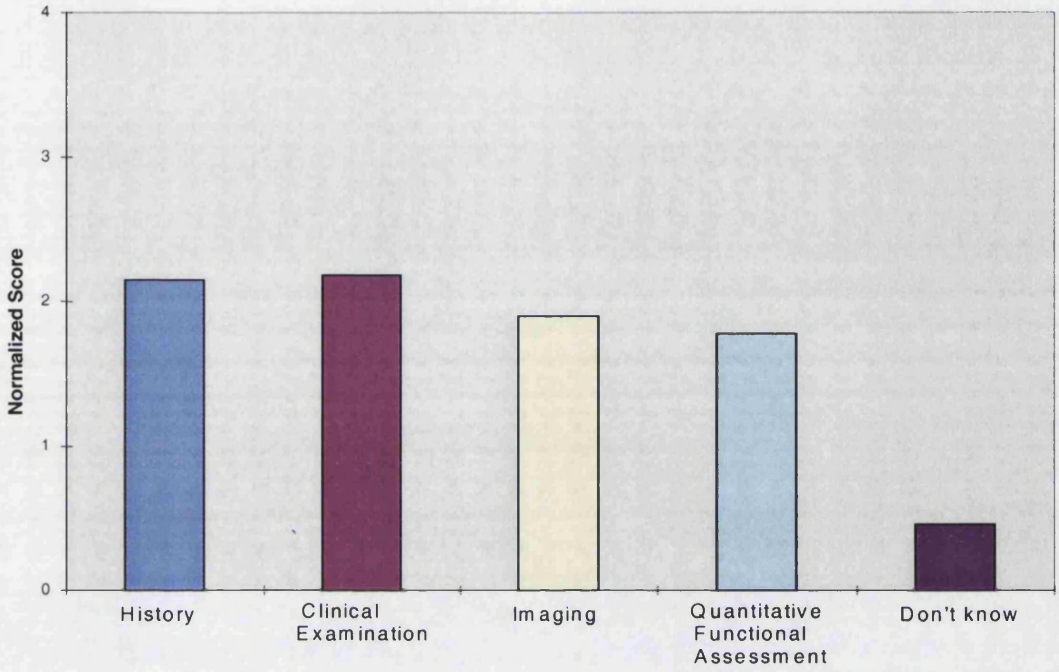


Figure 3.3 Histogram shows that clinical examination and quantitative functional assessment were ranked almost as important as history and imaging.

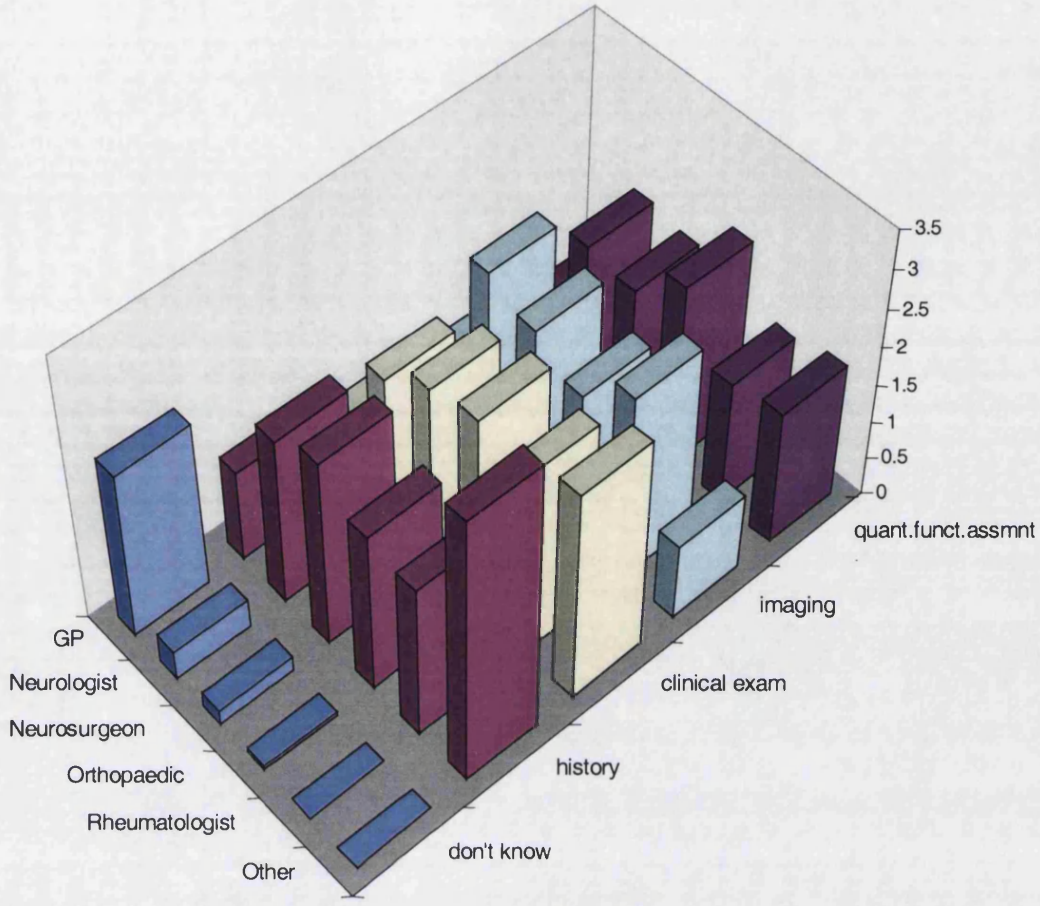


Figure 3.4 This 3D histogram shows inter speciality differences in the attitudes towards importance of different factors of assessment. GP's were unsure of CSM assessment.

3.4.2 *Open questioning- actual use of quantitative assessment and other criteria actually used. (Question 4).*

The actual use of quantitative assessment (i.e. practice as opposed to theory) was found to be very low, particularly when compared to the previous results indicating the relative importance of quantitative scales. Only 22 (19%) of clinicians actually used one or more assessment scales. GPs and Neurologists were found never to use scales. The Nurick scale was the one most commonly used (Table 3.1), but use of scales was in general characterised by infrequent use of a large number of different measures.

	Total	GP	Neurologist	Neurosurgeon	Orthopaedic Surgeons	Rheumatologist	Other
Nurick	6	0	0	2	4	0	0
FAM	4	0	0	1	0	1	2
FIM	4	0	0	0	2	1	1
JOA	4	0	0	1	3	0	0
SF36	4	0	0	1	2	1	0
Barthel Index	3	0	0	1	1	0	1
MDI	2	0	0	2	0	0	0
Walking Test	2	0	0	2	0	0	0
EMS	1	0	0	1	0	0	0
NDI	1	0	0	1	0	0	0
Button Test	1	0	0	1	0	0	0
QALY	1	0	0	0	1	0	0

Table 3.1 In general the scales were infrequently used although Nurick was the most commonly used scale.

The criterion actually used for assessment that was most commonly volunteered by clinicians was history, followed by overall function and then general examination (table 3.2) (question 7). There were differences between speciality that may reflect the natural bias of such specialities, e.g. no rheumatologists volunteered bladder function and 20% volunteered pain while 18% of neurologists volunteered bladder function and none pain. Disease progression over time was often highlighted by neurosurgeons and rheumatologists.

3.4.3 Attitudes to quantitative assessment:

Only 3 (2.6%) clinicians actually based their clinical decisions (e.g. whether to operate or to refer) on quantitative assessment scales, whilst 42 (36%) reported that assessment scales were useful for research/academic purposes (question 6).

The majority of clinicians (97.4% n=114) considered that there was no Gold Standard for assessment of CSM. The three that did consider there to be a Gold Standard cited different scales: the EMS, the Nurick and the Functional Assessment Measure (FAM) (question 3).

On rating the attribute of an ideal quantitative scale that would be most useful in actual practice (by summing the scores from 1 to 4 and dividing by the number of clinicians), "ease of use" was ranked highest at 1.65, followed by reproducibility (1.54), sensitivity to change (1.31) and finally validity (1.21). Note that this section was left blank by an appreciable number of respondents (question 5).

	Mean %	GP %	Neurologist %	Neuro-surgeon%	Orthopaedic Surgeon %	Rheumatologist %	Other %
Clinical examination	**44**	18	64	45	32	60	100
History	**58**	36	82	50	58	80	75
Overall function	**44**	27	36	50	42	60	75
Disease progression	**22**	9	18	41	11	40	0
MRI findings	**39**	9	55	59	32	40	0
Disability	**42**	18	36	41	53	40	75
Bladder function	**13**	9	18	14	11	0	25
Pain	**7**	9	0	9	5	20	0

Table 3.2 Criteria actually used for assessment by clinicians (% of each speciality that noted each parameter as being used by them in assessment). Some clinicians selected more than one criteria, hence total % is greater than 100%.

3.5 Discussion

This survey underlines the fact that current assessment of CSM patients is subjective and non-standardised (Lunsford, Bissonette et al., 1980b;Harsh, Sypert, Weinstein, Ross, and Wilson, 1987;Hadley and Sonntag, 1996;Sypert, 1992;Allen, 1952;Bakay, Cares, and Smith, 1970;Gonzales-Feria, 1975). There is a low level of use of standardised assessment scales and, when they are used, different scales are employed by different clinicians. These findings are likely to reflect a lack of confidence in the utility of quantitative assessment and a lack of agreement on which is best to use. This is also well illustrated by the almost universal acceptance that there is no "Gold Standard" for assessing CSM disease severity.

The closed questions asking the relative importance of certain measures are revealing in that they yield somewhat different results from those on actual practice. Possibly the former better reflect clinician's judgement of *ideal* practice. Certainly quantitative measures were dramatically underrepresented in actual practice compared with their relative importance (which was considered to be nearly as high as examination, history and imaging). Perhaps this reflects that quantitative scales are valued but currently existing scales do not meet clinicians' demands in terms of actual use. Another interesting finding is that many clinicians volunteered overall functional assessment and disability as a criterion they actually used on which to base clinical decisions, even though this is what quantitative scales purport to measure. This again highlights the difference between theory and practice. A related issue is the fact that relatively large numbers of clinicians felt that scales were useful for research purposes. Clearly, with poorly circumscribed selection criteria for surgery in CSM and a lack of quantitative assessment of initial severity, as well as of surgical outcome, any useful comparison of the efficacy of different management strategies and surgical techniques is difficult.

GPs and other non-surgical specialities were surveyed in this study because the attitudes of such clinicians may be an important factor in referral practices and speed of

referral. There was a complete lack of use of quantitative measures among these specialities (Table 3.1), even though in the UK most patients' treatment will depend initially on assessment by these clinicians.

It must be noted that certain clinicians' lack of focus upon quantitative measures and instead upon history, examination and imaging reflects that the latter are vital for initial diagnosis rather than just assessment of severity. Quantitative scales are only intended as an adjunct for rather than replacement for these practices. Thus the widespread use of scales *in isolation* by GPs would have dangers, because the primary role of such clinicians is in initial diagnostic assessment. On the other hand, quantitative scales have the advantage of assessment of *change* in individual patients and, once a diagnosis is made, non-surgical specialities such as GPs and neurologists are often well placed for monitoring of patients that do not immediately undergo surgery. The questionnaire does indeed identify that *disease progression*, i.e. measurement over time, is felt to be important, again highlighting the potential for a quantitative measure to be applied repeatedly and accurately on patients to monitor progress. In current practice, there are concerns that referrals are sometimes made too late, or made routinely rather than urgently.

Finally, it was instructive to note that, as well as the ability accurately to monitor disease progression, which would constitute sensitivity to change, closed questioning revealed that clinicians felt ease of use to be the most important attribute of a measurement scale, followed by reproducibility.

3.6 Conclusions

1. Quantitative measures are considered by clinicians to be important, yet are rarely used in actual practice. It would thus appear that clinicians recognise a need for quantitative assessment, even GPs and neurologists who were found never to use such scales, but that there is dissatisfaction with those scales currently available.
2. Given the current climate of evidence based practice, it would seem particularly timely to consider the development and application of new scales that better

reflect clinicians' attitudes and needs.

3. In designing a possible new scale, one should appreciate that a history of loss of function and the rate of progression were highlighted as being most important in determining management decisions, and that it was considered most important that a scale be easy to use and reproducible.

4. Preliminary Analysis of the use of Walking parameters in Assessing Severity of Cervical Spondylotic Myelopathy

4.1 Summary

To develop the use of a 30-m walking test as a quantifiable measure of severity of cervical spondylotic myelopathy (CSM) and a measure of the effects of decompressive surgical treatment, pre- and post-operative measurements were made of 30-m walking times and the number of steps taken over this distance in 40 CSM patients. The walking parameters were compared with a similar number of age and sex-matched healthy controls without myelopathy.

Both walking time and the number of steps taken were significantly worse in pre-operative patients than in control subjects. The walking data were highly reproducible over three trials. Post-operatively, there was a significant improvement in walking time and number of steps taken. The two walking parameters show a good correlation of change between walking time and steps taken, and are sensitive, reliable and easy to use.

4.2 Introduction

A criticism often levelled at functional scales used to assess neurological disease, that may be applied to CSM in particular, is that they lack the objectivity of a clinical neurological examination. For example, a scale based on patients' reporting of their own symptoms might be coloured by their general psyche and by how urgently they perceive they need treatment. It would be advantageous therefore to achieve further rationalisation of severity beyond the above-described scales of functional impairment.

The present study explores the use of a timed walking test (previously employed for other purposes) (Holden, Gill, Magliozzi, Nathan, and Piehl, 1984) as a more objective, quantitative and easy to use measure to quantify the severity of CSM both pre-

and post-operatively. Clinicians already tend to question patients on walking, as a "rough and ready" measure of the functional severity of spastic paraparesis and walking is of course an important aspect of the neurological examination that sensitively reflects long tract pathology. Thus timed walks may be an ideal means of bridging the gap between disability and impairment rating scales and, since clinicians are strongly swayed by their own clinical experience, a measure that may actually be readily accepted and used by practising clinicians.

4.3 Methods

4.3.1 *Subjects*

A total of 40 patients with CSM consecutively referred to the NHNN Neurosurgical Unit for consideration of decompressive surgery were studied. This patient group was compared with forty controls (normal subjects without myelopathy or other disability) and was matched for sex and, as far as possible, for age.

4.3.2 *Protocol*

Walking data for 30-m timed walks was collected in the manner described in Chapter 2 (Methods) both pre-operatively and also 2 to 6 months after surgical decompression. Each walking trial was repeated 3 times to test repeatability. Similar walking data were recorded from normal control subjects on a single occasion.

4.4 Results

There were 26 male patients of mean age 59.4 ± 11.9 (SD) years and 15 females of mean age 62.9 ± 18.2 (SD) years. In comparison, 26 male controls had a mean age 59.8 ± 12.5 years and 15 females had a mean age 61.9 ± 17.7 years. There was overall no

significant difference between the mean ages of the patients and controls (two tailed t-test on paired samples; p=0.65). Thus a good match was achieved, justifying further comparisons between patients and controls. The median length of hospital admission for surgery was 7 days and there were no peri-operative deaths.

4.4.1 Reproducibility

To assess the reproducibility of the walking time data, analysis of variance was performed as a two-factor (pre- and post-op) repeated measures test comparing the values for the three separate trials performed by each subject in each of the pre- and post-operative conditions. By far the greater variability existed between pre- and post-operative times (p=0.001) rather than between trials (p=0.995) (interaction; p=0.998), indicating that the walking data was highly reproducible and any major change in value was likely to represent a real change in the subject's functional ability. Thus only the mean of the three values was used in further analysis. The source of variation for number of steps taken again lay primarily between pre- and post-operative values (p=0.003) rather than between the three trials (p=0.981) and there was no interaction (p=0.959).

4.4.2 Controls

The mean pre-operative walking time over the three trials and over all the people in the control group was 24.3 ± 0.8 sec. (SE) (Both control and patient walking data were approximately normally distributed.) (Fig. 4.1). The mean number of steps taken to walk 30-m was 46.9 ± 1.2 steps (SE) (fig. 4.1).

4.4.3 Myelopathic patients

The mean pre-operative walking time for the CSM patients was 85.4 ± 11.2 sec.

(SE for 39 subjects). This was significantly worse than that for healthy controls (two-tailed t-test (unpaired, unequal variances);$p=2.4x10^{-6}$). The mean post-operative patient walking time was 64.7 ± 8.4 sec. (SE for 38 subjects) (fig. 4.1). There was a significant improvement in the patients following operation (two tailed paired t-test; p=0.0018). One patient was unable to walk this distance preoperatively but able postoperatively. Two patients were able to walk preoperatively but unable postoperatively. These patients' data were removed from both pre- and post-operative scores.

The mean number of steps taken to walk 30-m was 74.8 ± 5.3 steps (SE) preoperatively and 63.5 ± 4.2 steps (SE) postoperatively (fig. 4.1). The pre-operative number of steps was again significantly worse than that for controls (two tailed t-test; $p=5.4x10^{-6}$) and there was again a highly significant improvement in patients after the operation (two-tailed paired t-test; $p=5.87 x10^{-6}$).

It was clear from the control data that, despite the significant improvement after the operation, walking did not reach normal levels.

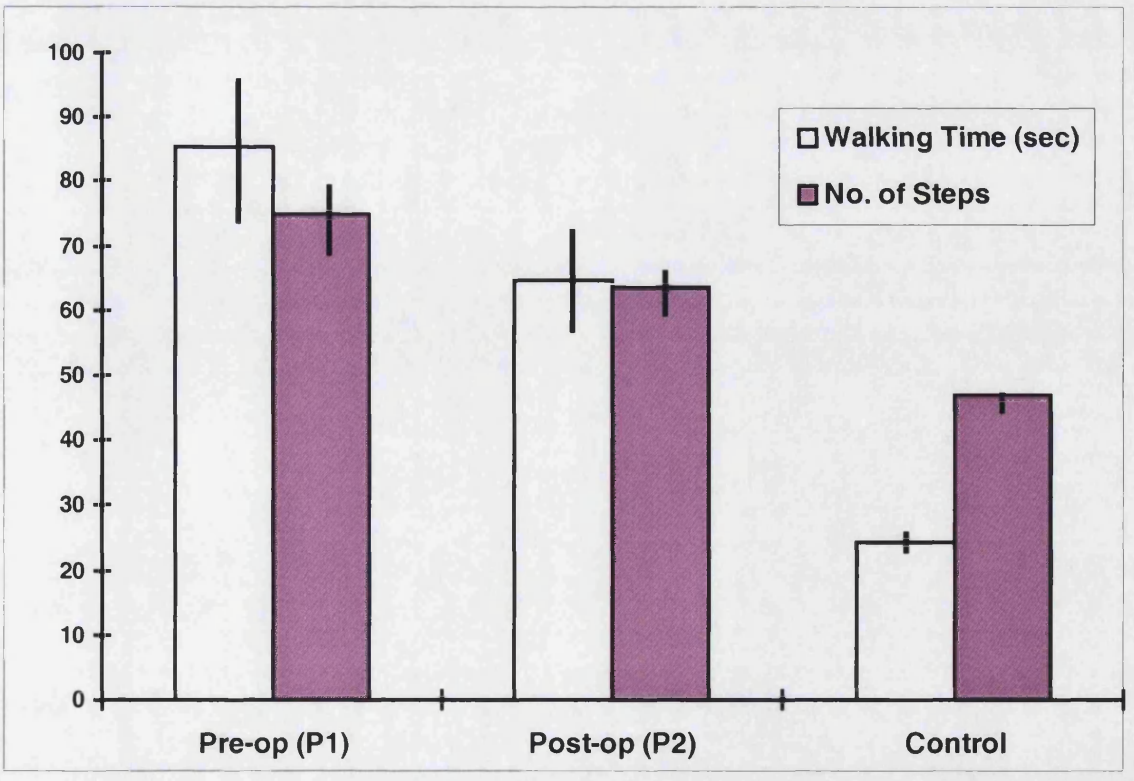


Figure 4.1: walking and number of steps taken in normal controls and patients pre-operatively (P1) and post-operatively (P2).

4.4.4 Sensitivity

The highly significant improvement following surgery indicates that the walking parameters have good sensitivity to change.

However, absolute sensitivity is also desirable; this is the sensitivity of a measure to distinguish different levels of disability. A measure of this is inter-subject variability (Steiner and Norman, 1996) and can be quantified by the coefficient of variation (standard deviation / mean). For pre-operative walking time this was 0.88 and for pre-operative steps taken 0.44. For controls the values were 0.21 for walking time and 0.15 for number of steps. The clear difference between pre-operative patients and controls is also a measure of sensitivity. (Note that rating scales cannot score controls.)

4.4.5 Relationship between walking time and number of steps

In patients both pre and post operatively the number of steps was around the same as the walking time in seconds so that on average patients took around a second for each step. In contrast, normal age-matched subjects took steps around twice as quickly. It was this extra increase in step speed that seemed to enable extra increases in overall speed towards normal, presumably because stride length was already near normal and further increases would clearly be limited by leg length. Alternatively, step speed may be impairment in myelopathy that responds less well to surgery.

4.4.6 Relationship between age and walking

It would be expected that age as well as the severity of disease might have an effect on the walking parameters. In the age-matched controls, this relationship was not found to be particularly strong (not illustrated) for number of steps taken (correlation coefficient=0.42) or for walking time (correlation coefficient=0.34). The corresponding

coefficients for preoperative patients are lower than for controls, as expected because there is also the influence of severity of disease (number of steps; correlation coefficient=0.32: time; correlation coefficient=0.26).

By linear regression analysis, the slope for steps in controls was 0.212 ± 0.07 steps/year (± standard error), that for steps in patients pre-op was 0.759 ± 0.37, that for walking time in controls was 0.118 ± 0.05 sec/year (fig. 4.2a) and that for walking time in patients pre-op 1.32 ± 0.79 sec/year (fig. 4.2b). Thus, walking time in patients had a steeper relationship with age than did steps taken, even though the strength of correlation (r-value) was greater for steps taken. Similarly, the slope on the regression line is much steeper for patients than for controls, even though the correlation was weaker. Perhaps this may be explained by the fact that age is the major factor affecting walking time in controls but the level of disability much more dramatically affects patients' walking. Nevertheless, despite the *relative* contribution of age in patients being lower, its *amount* of contribution in seconds per year is greater. This probably reflects that the effect of cervical spondylosis on disability is greater in more elderly patients, because there may be less "neurological reserve". Also, older patients may have worse disease than younger patients at presentation. The influence of age in patients might therefore be difficult to extrapolate from that in controls.

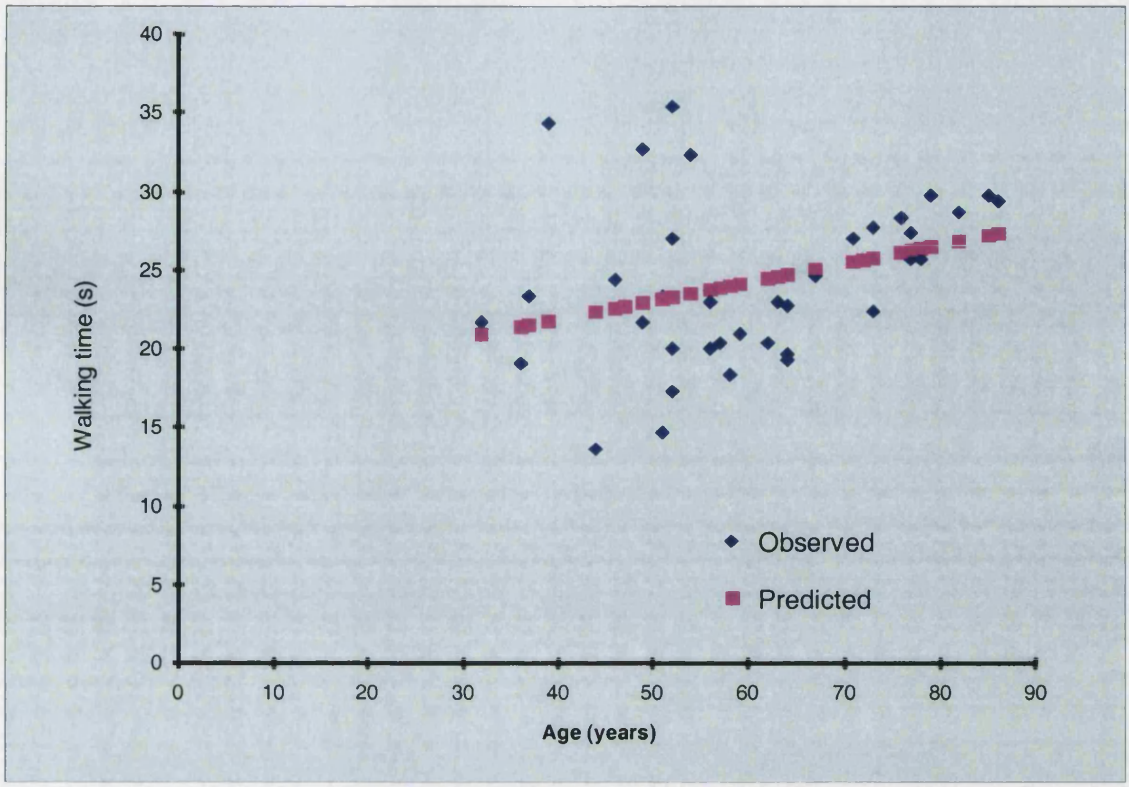


Figure 4.2a. Linear regression slope for walking time with age in controls

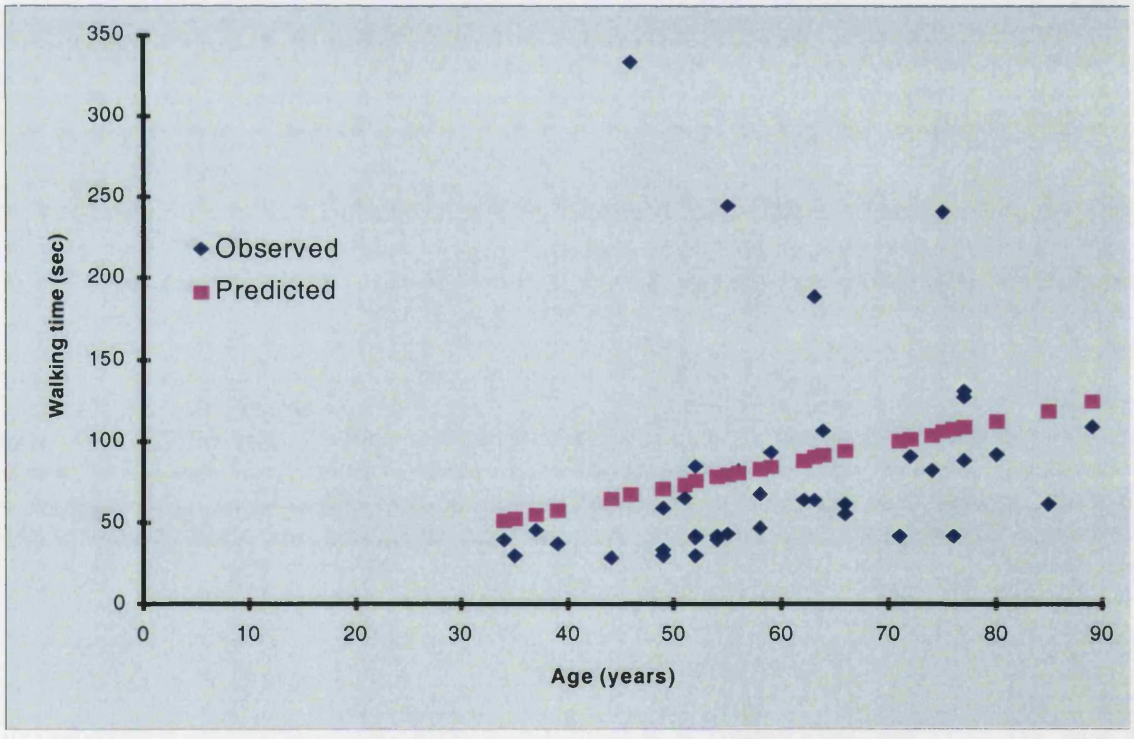


Figure 4.2b. Linear regression slope for walking time with age in patients pre-op.

4.5 Discussion

This study indicates that the easily performed walking test may be a suitable measure of severity of CSM. The test is shown to be reproducible and reliable, since there is a very low inter-trial variability (Steiner and Norman, 1996). The distance of 30-m is long enough to measure time relatively accurately and the nature of such an objective quantitative measure guarantees a high level of reliability.

The walking test is also a truly quantitative measure, i.e. a ratio measure where the numbers represent real values rather than abstractions (Dun-Rankin, 1983). True parametric correlations, comparisons with control data and correction for "normal" control variability are therefore possible. The ability to correct for the effect of age on normal walking time could allow extra sensitivity for low levels of disability that other scales would simply categorise as "normal".

A high degree of inter-subject variability is desirable for a scale that will distinguish patients' disease severity (Steiner and Norman, 1996). The ratio of pre-op standard deviation / mean (i.e. coefficient of variation) for walking time (0.88) was greater than for steps (0.44), suggesting that walking time is potentially the better measure. However, sensitivity is more than just variability; the variation must reflect real differences in disability. Thus comparison must also be made with the coefficients of variation of the controls that are all 'normal'. The values were 0.21 for walking time and 0.15 for number of steps. Thus walking time still shows the better variability even when accounting for the greater 'normal' walking time variability. In addition, it shows less strong correlation with age, a potential confounding factor. However, the slope of this variation is steeper.

The overall sensitivity to change is well demonstrated for the walking time data which show a dramatic improvement following surgery and would clearly seem to go

beyond the patient's self-assessment of 'being able to walk outdoors' or 'able to climb 5 steps' as is tested in the MDI. Furthermore, the more quantitative nature of walking data allows more detailed comparison with control data (on other scales all controls are scored the same – normal). It becomes clear that, despite the significant improvement after the operation, walking did not reach normal levels. Patients both pre and post operatively took on average around a second each for a step. In contrast, normal age-matched subjects took steps around twice as quickly. It was this extra increase in step speed that seemed to characterise normal walking. The slowness of walking cadence in CSM could represent a relatively irreversible component of spasticity.

Relevance and validity are important factors that will give clinicians confidence in using an assessment scale. Since walking time (incorporating maximum speed and endurance) is both an important aspect of patients' function and a sensitive measure of long-tract dysfunction as part of neurological examination, the test would empirically seem to be highly relevant to CSM severity and is obviously a true measure of walking. However, one must be alert to the fact that walking can be impaired by other common pathologies.

In conclusion, these preliminary data indicate that the timed walk may be highly suited to the assessment of severity of CSM. It is hoped that walking tests may prove useful as a practical aid for clinicians in pre-operative assessment and as a measurement tool in future trials on CSM operative techniques and management strategies. The next step in validating such use of walking parameters is to compare them directly with existing scales of CSM severity.

5. Comparison of Walking Parameters with Seven existing Scales used to Quantify Severity of Cervical Spondylotic Myelopathy and Post-operative Improvement

5.1 Summary

Difficulties in determining those patients with cervical spondylotic myelopathy (CSM) most likely to benefit from decompressive surgery might be addressed by accurate quantification of CSM severity as part of a trial determining the outcome of surgery. This chapter describes a study comparing the applicability of the walking parameters just described with various existing quantitative severity scales used in the measurement of CSM severity and in the measurement of the effects of surgical decompression. In addition to walking data, scores on the following scales were determined in 100 patients with CSM pre-operatively and then again six months following surgical decompression: Odom's Criteria, Nurick grade, Ranawat grade, Myelopathy Disability Index (MDI), Japanese Orthopaedic Association (JOA) score, European Myelopathy Score (EMS) and Short Form-36 Health Survey.

The walking parameters again recorded a significant improvement following surgery. All the scales did likewise. However, each had differing qualities of reliability, validity and responsiveness that made them more or less suitable. The MDI showed the greatest sensitivity between different severity levels, sensitivity to operative change and internal consistency. However, analysis of all the questionnaire scales into components that looked at different aspects of function revealed potential problems with redundancy and a lack of consistency when compared with each other or with the walking parameters.

This study provides a rational basis for determining the advantages and disadvantages of different existing scales in measurement of CSM severity and for making adaptations to develop a scale more specifically suited to a comprehensive surgical trial.

5.2 Introduction

The selection of appropriate patients for CSM decompressive surgery and the determination of the correct stage in the disease to operate remains uncertain. In fact (Rowland, 1992), has questioned the fact that surgery has any role in cervical spondylotic myelopathy, arguing that there has been no large prospective surgical series. While the lack of such data does not invalidate operative treatment, different clinicians do appear to vary greatly in their selection practices for decompressive surgery. The increasing demand for scientific justification of clinical practice certainly makes some form of large prospective comparison of different measures of assessment important.

The goal of the study described in this chapter was to provide such a prospective comparison by the measurement of walking parameters in a large group of 100 patients pre-operatively and then 6 months post-operatively. As described in chapter 2, a variety of quantitative assessment scales now exist that have or could potentially be applied to the quantification of CSM severity and so the scores on many of these scales were recorded at the same time to provide maximum impact and acceptance of the findings by various different groups who all tend to use different scales, to compare the applicability of these various impairment, disability and handicap scales to CSM patients and to compare such scales with the walking parameters as an attempt to validate the latter. If no one scale or parameter is found to be ideal, the study may at least help to determine the best and the worst features of the different measures in an effort to develop an ideal measure.

5.3 Methods

5.3.1 *Subjects*

One hundred patients with CSM were prospectively studied. They had been consecutively referred and accepted for decompressive surgery to the Neurosurgical Unit at NHNN. The median age of the patients was 58 years; there were 62 males and 38 females. All patients had the diagnosis corroborated by MRI and none had undergone previous neck surgery or had any other pathology that might have resulted in functional impairment. Of the 100 patients, 50 anterior cervical discectomies (Cloward or Smith

Robinson) and 50 posterior decompressions (laminectomies n=16, laminoplasties n=34) were performed by 6 different neurosurgeons.

5.3.2 Study design and data analysis:

Each patient was assessed by the same assessor. Walking parameters were measured as previously described.

In addition, scores for the following functional assessment scales were determined shortly before surgery and then again 6 months after surgery.

1. Myelopathy Disability Index (MDI): a disability scale applied to assessment of rheumatoid myelopathy and constituting a shortened form of the Health Assessment Questionnaire (HAQ), which in turn is adapted from the Activities of daily living (ADL) scale. Scores range from 0 (normal) to 33 (worst) (Casey, Bland, and Crockard, 1996).
2. Japanese Orthopaedic Association Score (JOA): a disability scale that attempts to look at various impairment categories such as disability related to upper motor neurone, radicular and sphincter deficits. Scores range from 0 (worst) to 17 (normal) (Hirabayashi, Watanabe, Wakano, Suzuki, Satomi, and Ishii, 1983;, 1994).
3. European Myelopathy Score (EMS): a scale adapted from the JOA for Western use that also includes pain assessment. Scores range from 5 (worst) to 18 (normal) (Herdman, Linzbach, Krzan, and et al, 1994).
4. Nurick score: a simple scale mainly focusing on walking disability, ranging from 1 (normal) to 5 (worst) (Nurick, 1972).
5. Ranawat score: a simple impairment scale, ranging from 1 (normal) to 4 (3B) (worst) (Ranawat, 1979).
6. Odom's criteria: a simple score looking at overall surgical outcome, ranging from 1 (best outcome) to 4 (no change or worse) (Odom, Finney, and Woodhall, 1958).
7. SF36 Health Survey: a complex health questionnaire measuring disability and handicap (% of normal 100%) (Ware and Sherbourne, 1992).

During post-operative assessment, the assessor was not shown the pre-operative scores.

The different outcomes measures were analysed with respect to their properties of internal consistency, sensitivity, validity and responsiveness (sensitivity to change), as outlined in chapter 3. Data were analysed statistically using the SPSS package version 9.

5.4 Results

5.4.1 *Patient and Operative Details*

The median length of hospital stay for the 100 patients was 8 days and there was a 3 % wound infection rate. There was one peri-operative death due to cardio-respiratory failure 3 weeks following surgery. Thus only 99 comparisons were available.

5.4.2 *Improvement following surgery*

Out of the 100 patients studied, 5 were unable to walk at all pre-operatively. Of these 5, one died 3 weeks post-operatively of cardiorespiratory failure, 3 became able to walk post-operatively and 1 remained unable to walk. Of the patients able to walk pre-operatively, all were also able to do so following surgery.

There was a highly significant improvement following surgery both in the walking time and in the number of steps taken ($p<0.001$; paired two-tailed t-test) (fig. 5.1). All the other scales also showed highly significant improvement following surgery ($p<0.001$; Wilcoxon signed rank test) (table 5.1, fig. 5.2 for SF36 categories, fig. 5.3). Note that in some of the scales (EMS, JOA, and SF36), an improvement is represented by a *higher* score and that Odom's criteria only record operative results so there are no pre- and post-operative values. There were a minority of patients who scored worse following surgery (eg 8 out of 99 for the MDI). On each scale, these were slightly different patients (see correlations section).

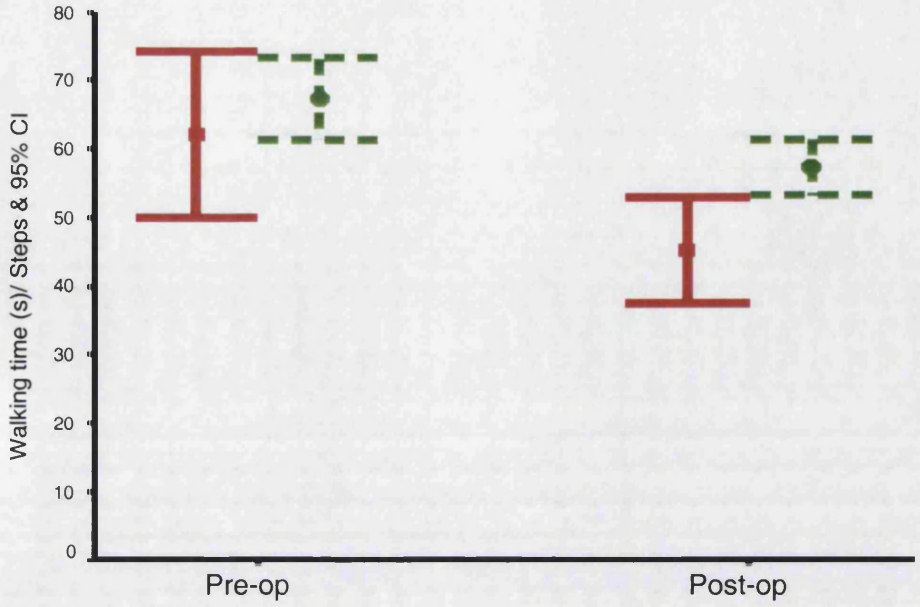


Figure 5.1. Means and confidence intervals of the patients' walking times (red bars) and number of steps taken (green hatched bars) pre-operatively and post-operatively.

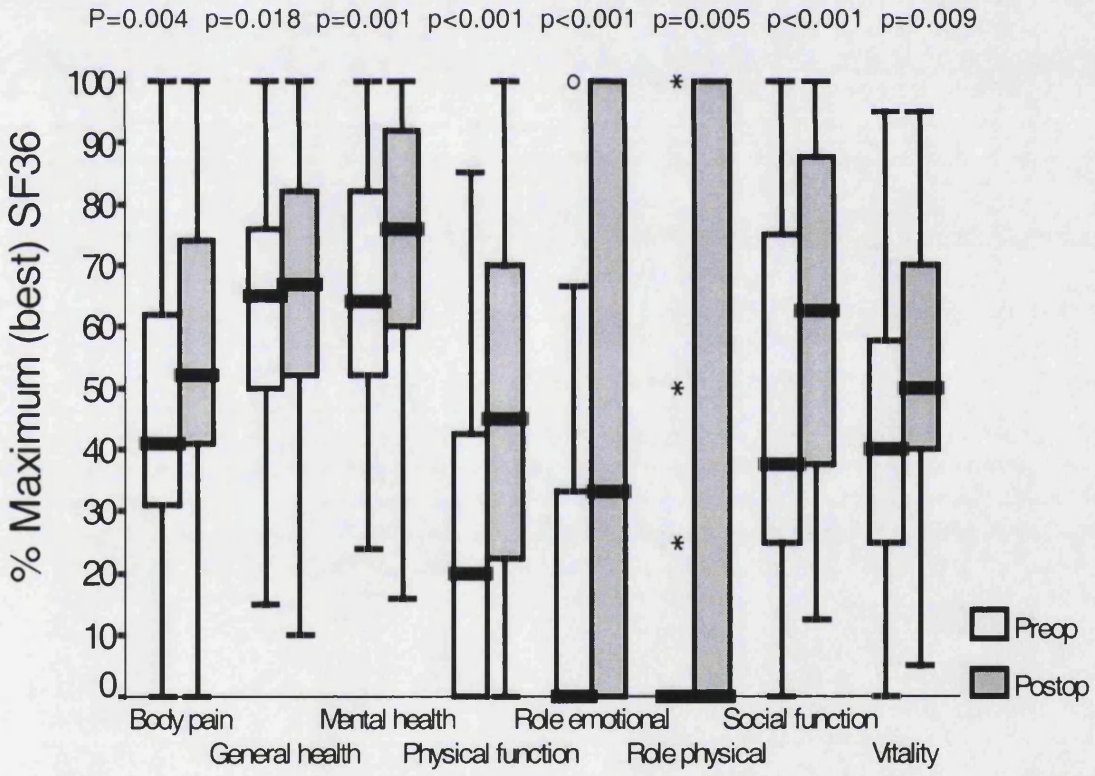


Figure 5.2. Box plots of pre and post-operative scores for the 8 categories of the SF-36 Questionnaire. These scores have all been transformed to percentages for comparison, where 100 % is the best possible score. Each category shows significant improvement following surgery (Wilcoxon).

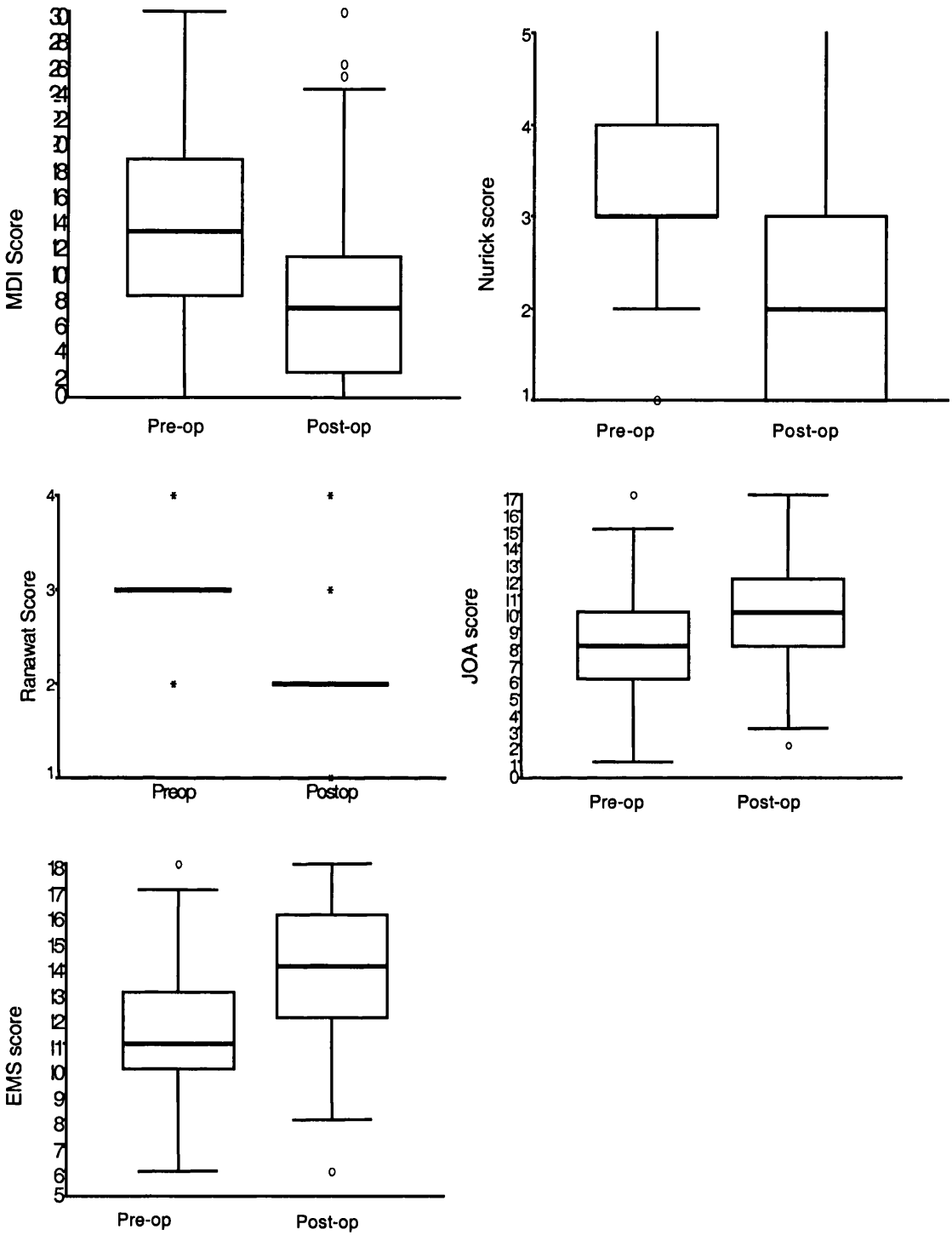


Fig. 5.3. Box and whisker plots of the 100 pre-operative and 99 post-operative scores of all the patients on 5 different scales. (One patient died shortly following surgery.) For the MDI, the Nurick and the Ranawat scales a better score is a lower value, while for the EMS and JOA better scores are represented by higher values. The circles represent outlying values between 1 ½ and 3 interquartile intervals and the stars represent extremes greater than 3 interquartile intervals. In all cases, the improvement following surgery was statistically significant (Wilcoxon) (table 6.1).

Scale	Signifi-cance of Improve-ment	Sensitivity to Change	Coefficient of variation Pre-op	Coefficient of variation Post-op	Internal Consistency (Cronbach,s ∝) Pre-op	Internal Consistency (Cronbach,s ∝) Post-op
Walking	P<0.001	0.38	0.91	0.8	0.85	0.83
Steps	P<0.001	0.18	0.37	0.33	0.85	0.83
MDI	P<0.001	0.52	0.85	1.29	0.92 (11 scores) 0.76 (4categories)	0.95 (11 scores) 0.81 (4 categories)
EMS	P<0.001	0.18	0.27	0.29	0.68	0.77
JOA	P<0.001	0.21	0.5	0.4	0.72 (6 scores) 0.66 (4categories)	0.73 (6 scores) 0.65 (4 categories)
Nurick	P<0.001	0.42	0.33	1	-	-
Ranawat	P<0.001	0.34	0	0	-	-
SF 36	P<0.001	0.32	0.41	0.68	0.82 (36 items)	0.86 (36 items)

Table 5.1 Comparison of properties of different scales. The significance of improvement is the p-value of the operative change. Sensitivity to change is mean of (pre-op score – post-op score)/ median of all scores. Coefficients of variation pre-op and post-op and the reliability (cronbach's α) pre-op and post-op are also shown for all scales.

5.4.3 Characteristics of the different measures

5.4.3.1 Sensitivity to change: The two walking parameters and the rating scales all had highly significant improvements following surgery. However, a desirable property of a test would also be its ability to show an important *magnitude* of change, indicating a high responsiveness or sensitivity to post-operative change. This property was quantified by calculating the "normalised change", which was the mean of the differences following surgery across all the patients divided by the overall median of the pre- and post-operative scores (table 5.1). (The median was chosen because most of the scales give non-parametric distributions of values, but the differences between pre- and post-operative values more closely followed a normal distribution and so the mean was calculated for this parameter.)

It was seen that walking time was better than number of steps taken in that it showed a greater post-operative change (table 6.1). Overall, the MDI score showed the greatest sensitivity to change. The EMS was the worst measure of change.

5.4.3.2 *Sensitivity*: As well as sensitivity to change, it is desirable for a test to record a broad range of absolute values across a population, indicating that the scale has a high absolute sensitivity in detecting differences in severity between different patients. A measure of this is the coefficient of variation of the data. For non-parametric data this is calculated by the interquartile range divided by the median value. This method was also used for the walking data to allow a direct comparison.

It was found that the walking time showed a greater sensitivity than the number of steps taken. The walking time was comparable with the MDI, the best of the other scales. The Ranawat score had poor sensitivity for distinguishing patients with different levels of severity, because the range across the patients was narrow. This is illustrated by the fact that the box plot shows a single horizontal line instead of a box (fig. 5.3). Nearly all pre-operative patients were scored at one level and post-operatively at a level one grade better, indicating that the Ranawat score nevertheless records a postoperative

improvement.

The pre-operative and post-operative coefficients of variability were generally comparable, apart from the Nurick test where the pre-operative coefficient was 0.33 and the post-operative coefficient 1.0. Since the patients generally improved post-operatively on all measures, this indicates that the Nurick test is relatively better (gives greater range of values) at distinguishing milder affected patients than more severely affected patients.

5.4.3.3 *Internal consistency*: If different questions in a multi-part questionnaire are attempting to measure the same parameter, then there should be consistent scoring within patients. This is measured by Cronbach's alpha, a normalised measure of correlations between multiple components of a scale. A score of 1 indicates a perfect correlation. (The SF36 internal consistency for the 36 items used converted scores before combination into 8 categories.)

Internal consistency can be construed as either potentially desirable or undesirable. A high value indicates reliability i.e. the scoring of the different aspects of a questionnaire is consistent. However, if the different aspects are intended to record different aspects of function, a high alpha score simply means that when one aspect e.g. bladder function is poor in myelopathy, other aspects e.g. Upper motor neurone (UMN) leg function are also poor in parallel, and this therefore suggests some redundancy. The MDI score gave a high internal consistency for the 11 questions (table 5.1), but it is possible that this simply reflected that the same aspect of function was being asked repeatedly or that different aspects always vary in parallel. When the 11 questions of the MDI were split into 4 categories (walking, hand function, transfers and dressing), the alpha scores were somewhat lower. This is appropriate, since if different questions within a questionnaire are designed to address different parameters, then it is not appropriate to seek high internal consistency.

The walking time and number of steps taken had a high level of consistency with each other, again suggesting some redundancy.

5.4.3.4 *Comparisons between different scales*: To explore the validity of the different scales, correlation coefficients were calculated for the pre-operative scores (table 5.2A), post-operative scores (table 5.2B) and for the changes following surgery (table 5.2C). All correlations were corrected for the fact that some scales recorded no disability as the maximum value, while others recorded no disability as the minimum value.

All the scales showed significant pre-and post-operative correlations (Spearman's rank correlation at the p=0.01 level) (table 5.2A, 5.2B). Some scales were correlated better than others; the best correlation was found post-operatively between the MDI and the EMS scales (r= 0.82) which are both disability questionnaires, while the poorest correlation was post-operatively between the overall averaged SF36 (measuring handicap and disability) and the Ranawat (measuring neurological impairment). (Note that it is unclear that averaging all 8 SF36 components into an amalgamated percentage has been validated.)

The correlations were poorer when comparing operative changes. Many values were close to zero or even negative (table 5.2C). The best correlation of change between pre-existing scales was for the Ranawat and Nurick scales (r=0.55, Spearman rank correlation), which indicates that they scored change in a similar way. However, by far the best correlation overall was between change on walking time and change in number of steps taken (0.75).

The correlations with the Odoms' criteria for success of surgery were poor, partly because the latter score does not really distinguish between a patient who improved well following surgery from one who simply remained well following surgery.

A	Pre-op MDI	Pre-op EMS	Pre-op RANAWAT	Pre-op NURICK	Pre-op JOA	Pre-op SF36	Pre-op TIME	Pre-op STEPS
Pre-op MDI	1	-	-	-	-	-	-	-
Pre-op EMS	0.75	1	-	-	-	-	-	-
Pre-op RANAWAT	0.51	0.61	1	-	-	-	-	-
Pre-op NURICK	0.66	0.69	0.71	1	-	-	-	-
Pre-op JOA	0.56	0.62	0.47	0.59	1	-	-	-
Pre-op SF36™	0.48	0.42	0.31	0.38	0.40	1	-	-
Pre-op TIME	0.58	0.45	0.25	0.57	0.37	0.36	1	-
Pre-op STEPS	0.52	0.43	0.28	0.53	0.38	0.41	0.81	1

B	Post-op MDI	Post-op EMS	Post-op RANAWAT	Post-op NURICK	Post-op JOA	Post-op SF36	Post-op TIME	Post-op STEPS
Post-op MDI	1	-	-	-	-	-	-	-
Post-op EMS	0.82	1	-	-	-	-	-	-
Post-op RANAWAT	0.67	0.63	1	-	-	-	-	-
Post-op NURICK	0.71	0.74	0.75	1	-	-	-	-
Post-op JOA	0.57	0.72	0.42	0.51	1	-	-	-
Post-op SF36™	0.35	0.35	0.25	0.36	0.37	1	-	-
Post-op TIME	0.44	0.51	0.38	0.42	0.54	0.43	1	-
Post-op STEPS	0.44	0.47	0.43	0.44	0.48	0.39	0.82	1

C	MDI Change	EMS Change	RANAWAT Change	NURICK Change	JOA Change	SF36 Change	ODOM'S Change	TIME Change	STEPS Change
MDI Change	1	-	-	-	-	-	-		
EMS Change	0.27	1	-	-	-	-			
RANAWAT Change	0.22	0.23	1	-	-	-	-		
NURICK Change	0.32	0.32	0.55	1	-	-	-		
JOA Change	0.15	0.35	0.02	0.19	1	-	-		
SF36™ Change	0.22	0.12	0.003	0.13	0.28	1	-		
ODOM'S Change	0.02	0.27	0.33	0.25	0.24	0.19	1		
TIME Change	0.32	0.13	0.07	0.27	-0.10	0.15	0.09	1	
STEPS Change	0.19	0.13	0.01	0.23	-0.07	0.23	0	0.75	1

Table.5.2 A, B, C. Correlation coefficients between the different pre-operative scale scores (5.2A) and the different post-operative scale scores (5.2B) and correlation coefficients between operative changes (i.e. pre-operative minus post-operative scores) recorded by the different scales (5.2C).

5.4.3.5 *Breaking down Scales into Components*

The generally poor correlation between scales, with better correlation between more similar scales (e.g. the postoperative MDI and EMS scores) could be due to some scales measuring different aspects of function or impairment. This was initially investigated by empirically dividing the multi-part scales into components measuring certain aspects of disability or impairment. This breakdown might also reveal that different individual aspects have different potentials for improvement following surgery. Thus, the Normalised Changes, measuring the magnitude of operative change (sensitivity to change) of the different components of the three multi-part disability questionnaires were calculated and compared (table 5.3).

A reasonably consistent trend was apparent across the scales revealing that good improvement tended to occur in hand function as assessed by all three scales addressing this aspect, while both scales looking at sphincter function showed that it remained little changed by surgery. Within the SF 36, physical and social function and social role changed most (fig.5.3), but no corroboration was available for these parameters since they were not measured by any other scale. The findings nevertheless in general support the possibility that the poor correlations might be better if one compared specific aspects of CSM rather than overall scales. However, since the scale components have not been validated when looked at individually, one has to interpret differences in improvement between these specific aspects with caution. For example, the greater improvement in hand function after surgery might simply reflect a greater sensitivity of the questionnaires to this component rather than a genuinely greater improvement.

	Walking	Hand	Dressing	Sphincter	Washing/ Transfers	Pain	Sensory Loss
MDI sensitivity to Change	0.58	0.70	0.35	-	0.42	-	-
EMS sensitivity to Change	0.2	0.22	0.2	0.03	-	0.22	-
JOA sensitivity to Change	0.21	0.35	-	0.04	-	-	0.33

Table.5.3 Three scales were broken down into their component aspects and sensitivities to change recalculated for these separate components. For example, the JOA has questions relating to walking, hand and sphincter function and sensory change. The hand function components recorded by these scales change much more than bladder-related components.

5.4.3.6 Correlations of Components:

In order to seek some validation of the component sensitivities and to explore why the overall scale correlations of operative change were low, the next step was to perform correlations between these components in a similar way to the correlations performed above for the overall scales. Thus, the components of the multi-part scales questioning walking function were directly correlated with each other as well as with the Ranawat and Nurick scales (which have a one-dimensional measure primarily based on walking), while hand and bladder components were similarly correlated between those scales that had aspects pertaining to these components (table 5.4A, B, C).

It was found that, particularly for hand and bladder function improvement, correlations were still very poor. The correlation of operative changes for two apparently similar questions on the JOA and EMS, namely bladder function, was only 0.23. On analysing individual patient's responses, the inconsistencies were clear. For example, patient number 10 indicated his bladder became worse post-operatively on the EMS, going from normal to inadequate, but on the JOA he reported only a mild disturbance both pre- and post-operatively.

Comparing walking components with the timed walk parameters, (table 5.1), correlations of operative changes were if anything worse than correlations for the overall scales! However, the EMS walking question and JOA walking question operative changes did correlate better (0.48) than the respective overall scale change.

A	MDI Walk Change	EMS Walk Change	RANAWAT Change	NURICK Change	JOA Walk Change	TIME Change	STEPS Change
MDI Walk Change	1	-	-	-	-	-	-
EMS Walk Change	0.07	1	-	-	-	-	-
RANAWAT Change	0.26	0.25	1	-	-	-	-
NURICK Change	0.34	0.23	0.55	1	-	-	-
JOA Walk Change	0.13	0.48	0.19	0.29	1	-	-
TIME Change	0.20	0.10	See above	See above	0.16	1	-
STEPS Change	0.14	0.13	See above	See above	0.16	See above	1

B	JOA Hand Change	MDI Hand Change	EMS Hand Change
JOA Hand Change	1	-	-
MDI Hand Change	0.12	1	-
EMS Hand Change	0.25	0.26	1

C	EMS Bladder Change	JOA Bladder Change
EMS Bladder Change	1	-
JOA Bladder Change	0.23	1

Table.5.4 A, B, C. Correlation coefficients between operative changes recorded by different scales where the scales have been divided into their walking (A), hand function (B) and bladder function (C) components. Splitting the scales and comparing their common functional components seems not to improve the correlation coefficients.

5.5 Discussion

All the quantitative measures of CSM severity satisfied the most basic requirement of a scale useful in assessing the effects of surgery in that they were all able to demonstrate a significant improvement in score following surgery. This consistent finding is of course also indicative of a genuine benefit resulting from such intervention. However, a true and meaningful effect of surgery would be better demonstrated by a study that included a period of follow up longer than 6 months and one that included a comparison with a similar group of CSM patients that were not operated upon. The primary purpose of this study was instead to investigate the properties of existing scales and compare them with the newly developed walking parameter tests.

Sensitivities of Different Scales

While all the scales showed significant improvement following surgery, they have other properties that make them more or less suitable for assessment of CSM. The MDI is sensitive to change and also gives a wide range of absolute values, which means there is good sensitivity to differences between patients. On the other hand, the Ranawat score, while being sensitive to change, was very poor at distinguishing different levels of absolute severity and so would be completely unsuitable for a study looking at differences between patients. This study, in looking at both pre- and post-operative scores, thus illustrates the important point that it is insufficient to attempt validation of scales only on absolute measurements; their properties may be considerably different if the scales are also to be used to assess the effect of operative or other interventions. In addition, widely differing absolute sensitivities between pre- and post-operative measurements suggests that different scales may have different applicability to different patient groups. For example, the Nurick score had a much greater sensitivity post-operatively, suggesting a greater ability to distinguish between different levels of severity at the milder end of the scale.

The two walking parameters compared favourably with most of the other tests, with the walking time being superior to the number of steps taken. The fact that walking

data is a ratio measure as opposed to an ordinal scale such as the MDI means that a given proportional change is statistically much more powerful.

Internal Consistency of Different Scales

The multi-part questionnaires had good internal consistency (internal reliability), particularly the MDI, suggesting that the questionnaires were being reliably completed. However, the high level of reliability may entail some redundancy where very similar questions concentrating on the same aspect of disability are asked repeatedly. Even worse, if different aspects of dysfunction are considered sometimes to be affected to different degrees in different patients, it would seem inappropriate that questions testing these different aspects always score too similarly. The lower alpha score when the MDI is divided into categories comparing different aspects does suggest some genuine effect in distinguishing these categories. Nevertheless, the presence of multiple questions within the same category, while not resulting in poorer sensitivity and sensitivity to change, does point to redundancy and therefore inefficiency.

A glance at the questions of the MDI (appendix 9) reveals that it tends to ask repeated questions on a few limited categories of disability. After the initial demonstration of high internal consistency during an initial study, indicating that the patients answer the questions reliably, perhaps redundant questions could simply be removed when designing an ideal scale used in assessing CSM severity. The high alpha score between walking time and number of steps again means that measuring both may be unnecessary. Since the walking time showed better sensitivity and sensitivity to change, perhaps counting the number of steps taken could be dispensed with in the future, making the test even more convenient to perform.

Intra-rater and inter-rater reliability were not investigated in this study. The MDI, EMS, JOA and SF 36 are patient rated, and so inter-rater reliability may be irrelevant for such scales. Instead, internal consistency is a measure of reliability across questions within the questionnaire. The Ranawat and Nurick scores are simple and one-dimensional and have previously been shown to have good reliability. Regarding the walking

parameters, using a stopwatch and counting steps are clearly objective and reliable in themselves.

Validity – Correlations between Scales

The concept of internal consistency does not necessarily imply validity and accuracy, i.e. whether or not a scale is actually measuring what it purports to measure (Wassertheil-Smoller, 1995). Scales are ideally validated by correlating them with a gold standard (Bland and Altman, 1986). This is most relevant when they are used as a convenient surrogate for a gold standard definitive investigation that is invasive, risky or cumbersome, or perhaps when used to predict an outcome that eventually becomes clear over time. Although such previously validated tests might be considered to be the "gold standard", this only applies to the population type for which the test was validated. For example, the SF36 cannot be said to be thoroughly validated in the context of this study because it has not previously been applied specifically to peri-operative CSM. Chapter 4 has already indicated that there is no existing agreed "gold standard" for CSM measurement. Thus, the scales were simply correlated with each other to see if certain inconsistencies became apparent.

It was found that, while correlations between similar scales were sometimes high, correlations between recorded operative changes were poor. This is because change is likely to be a much more sensitive indicator of dissimilarities between scales. For example, if a patient generally scores well on different scales pre-operatively and there is only a small post-operative improvement, the changes may well be in different directions on the different scales, while the post-operative absolute scores all still remain generally high. These highlighted differences between scales could reflect aspects of change that some scales measure which others ignore. Thus, a mildly affected patient may generally score quite highly, but operative decompression might change certain aspects much more than others. This point again illustrates the importance of validating scales by looking at changes rather than confining assessment to patients in the static state.

A notable exception to the poor operative change correlations are the two walking parameters, reflecting that walking numbers are true ratio measures which stand up to comparisons of proportions at all points in the scale.

Breakdown of Scales into Components

Subdividing the scales on empirical grounds into different functional components revealed differences between components, with hand function showing the greatest improvement, walking showing moderate improvement and bladder function showing minimal improvement. This might suggest that a quantifiable hand function task analogous to the walking test, such as a timed 9-hole peg task, might also be a responsive measure of CSM. The two might perhaps be combined to give an overall "upper and lower limb myelopathy (and possibly also radiculopathy) score".

However, these results must be interpreted with caution since they could reflect that different scales are simply better at measuring changes in different aspects of function rather than there being real differences in change of function. Indeed, when one actually correlates these different aspects of function by correlation of the components between the scales, the coefficients are often no better than for the overall scales. This also applied on comparing walking components with the walking parameters, which must surely be considered the "gold standard" of walking performance. Such findings throw doubt upon the validity of making strong inferences about the separate components of a scale and most importantly suggests that the poor overall correlations of improvement between the overall scales cannot be explained on the basis that the different scales record different aspects of this improvement, *but instead lead one to question the validity of some or all of the scales.* On review of individual patients' responses, it is clear that apparently similar single questions are sometimes answered very differently in different scales, possibly due to the phrasing of such questions. While it may be argued that a walking time is only relevant for the patient walking 30-m, it is at least a valid measure of this particular function. Given the poor correlation of walking components of scales with

each other and with walking time, it is unclear for example what an improvement in MDI actually means for the patient.

Another important precept of a multi-part scale is that there is an overall unidimensionality i.e. "overall severity". Thus the scale simply adds all the different components from which patients with myelopathy might suffer. No hierarchy of components is considered at all, other than perhaps more questions being asked on areas that are more important for patient functioning. This study has addressed the relationship between the components of different scales and found that, particularly when looking at *changes* in severity, this unidimensionality cannot be applied – some components deteriorate while others improve and there is no consideration of which are more important. While including multi-dimensional components might pick up disparate aspects of disability, statistical analysis on the score as a whole may be invalidated.

Such problems do not of course arise for a simple timed walk. The fact that the walking time's sensitivity and sensitivity to change compared favourably with those of most of the multi-part questionnaires indicates that the walking test does not seem to suffer through focussing on one aspect of function. Moreover, the poor correlations of similar aspects of function as measured by the different questionnaires only seem to lead to problems with interpretation and validity.

Conclusions

An ideal scale should be as quantitative as possible and show good sensitivity between patients and sensitivity to change. It should also be scored reliably and be simple to use. Of the pre-existing scales investigated, the MDI best reflects these characteristics. This scale constitutes a questionnaire that focuses upon a limited range of aspects of disability; the findings indicate that such a scale does not necessarily suffer in terms of sensitivity. Instead, repeated questioning on similar aspects of function may reflect redundancy. Moreover, the poor correlations between the operative changes recorded by the overall scales and their components indicates that repeated questions on different or even similar aspects of function may actually reveal considerable inconsistencies. There

is no increased validity to be gained from a scale combining these different aspects. In addition, there is often a lower rather than greater sensitivity to change in severity or to different severity levels.

Conversely, a scale measuring a limited aspect of dysfunction is likely to have a greater reliability. The most limited scale, namely the walking test, has a clear advantage arising from its objectivity and inherent reliability and also benefits, as seen by the strong correlation of change in walking time with the change in steps taken, in being a ratio rather than an interval or ordinal scale. Moreover, it has been demonstrated not to suffer in terms of sensitivity or sensitivity to change. Finally, correlation between walking time and the Nurick scale (a functional rating of walking alone) was no stronger than with the MDI (which measures overall function), indicating that the walking test might be used as a valid stand-alone marker for overall myelopathic severity.

6. Use of Walking data in Assessing Operative Results for Cervical Spondylotic Myelopathy – Long Term Follow Up and Comparison with Controls

6.1 Summary

In this study, the role of surgical decompression in CSM is explored more fully by using the walking test and other measures on a group of patients to assess severity pre-operatively and at regular intervals over the following three years. Such patients are compared with a matched group of control patients who have myelopathy but who remain unoperated upon.

The results, both from walking data and from the other scales, indicate a lasting benefit from surgery extending three years beyond surgery. Patients would appear to gain a significant recovery of function and then stabilise. In contrast, unoperated patients suffer an ongoing deterioration. Other points brought out, particularly from the walking parameter data, are that age and greater initial disease severity may increase rather than lessen the benefits of decompressive surgery.

6.2 Introduction

The previous chapters have indicated that the 30-m walking test is suitable for the measurement of CSM severity and shown that it compares favourably with existing measurement scales on a number of points. The goal of this chapter is to go beyond validation of the walking test and actually use it to answer important questions about the natural history of CSM and the effects of surgical treatment.

The need for a determination of the effectiveness of surgery for CSM was identified as long ago as the middle of the last century, when surgeons performing the first decompressive procedures appreciated their unpredictable benefits and risks of morbidity. In one of the earliest studies (Odom, Finney, and Woodhall, 1958) developed a simple four point scale of operative outcome and used this to demonstrate an "excellent" result in 94 out of 175 cases. However, it is unclear how many of these patients actually

had myelopathy rather than just cervical radicular problems (Robinson, Walker et al., 1962). adopted a similar scale which was later applied to show a "good to excellent" result of anterior cervical fusion in 53% of cases of myelopathy (Andrews, Gentchos et al., 1971).

Of course, a measure considering only operative outcome cannot be used to look at absolute severity or to make a comparison with unoperated cases. The importance of looking at these other factors to determine the real effects of surgery upon the natural history of CSM was also identified when management of CSM was still in its relative infancy. The first such study by Lees in 1963 (Lees and Turner, 1963) concluded that CSM was a rather benign condition with long periods of non-progression. Later, Nurick (Nurick, 1972) applied his own walking-biased severity (not outcome) scale to a large number of cases in the literature at that time and considered that laminectomy was of benefit mainly only in milder cases. He did consider surgery more appropriate for older patients because he discovered that such patients tended to deteriorate more rapidly if left untreated. However, a more recent re-analysis of these data have indicated a more severe natural deterioration than was originally reported and a greater relative benefit of surgical intervention in all age groups (Ball and Saunders, 1992).

One of the most commonly used severity/outcome measures has been the Japanese Orthopaedic Association score. This multi-aspect score has been applied in a number of studies, including those directed at determining which case characteristics predict a good surgical outcome. A formula comparing pre- and post-operative scores gives the recovery %. Fujiwara and his colleagues (Fujiwara, Yonenobu et al., 1989) reported that transverse cord area on pre-operative imaging was the most important predictor of good JOA recovery, where patients with a relatively unshrunken cord had a better prognosis. Younger patients were also found to do better following surgery. More recent corollaries of this study include a report that re-expansion of the cord on a post-operative axial MRI follow-up scan correlated with JOA improvement (Baba, Maezawa et al., 1997) and a study of 27 patients reassessed after a variable interval (12 to 96 months) that revealed JOA improvement only in a sub-group of patients below 60 years.

A criticism levelled at the JOA is that an important criterion is chopstick use, which might not be relevant to Western patients. In addition, many Japanese myelopathic patients suffer from ossified posterior longitudinal ligament (OPLL) pathology, which is very uncommon in Western Countries. The European Myelopathy Score (EMS) was developed to answer such difficulties (Herdman, Linzbach, Krzan, and et al, 1994) and was used to show that 60% of myelopathic patients had an unchanged or improved grade of severity following surgery. Unfortunately, studies using both the JOA and the EMS quote recovery rates when the formula allows no category to include patients who deteriorated following surgery! Thus recovery rates may be applicable only "for those who improved".

The EMS has not really become widely utilised and thus to date no one outcome measure has gained universal acceptance. In his review of the CSM literature up to the 1990's, Rowland (Rowland, 1992) was as a result only able to compare studies by "good" "fair" and "worsened" outcome. This objective review of disparate data resulted in his painting a rather pessimistic picture of the effects of surgery, where only a 50:50 chance of a good result could be put forward against a significant operative risk. This opinion has been challenged on the basis of factors such as diagnostic errors and lack of consideration of the particular procedure performed or the relative skill of the surgeon (Alexander, 1996). However, the point remains that the objective justification for the huge numbers of decompressive operations performed in modern practice remains tenuous.

A major deficiency in recent studies is the lack of comparison with a control unoperated population of CSM patients. Such studies were of course easier in the time of Lees 1963 (Lees and Turner, 1963) before decompression gained such wide acceptance. One recent study compared 20 decompressed with 23 unoperated patients and found that there was a significant gain in function and reduction in pain in the operated group and a deterioration in activities of daily living in the untreated group (Sampath, Bendebba et al., 1999). However, another 19 patients were lost to follow up. Furthermore, it was the surgeons themselves who were recommending patients go into the treated group on the basis of their myelopathic features.

Kadanka and colleagues looked at 48 patients who were randomised into surgery versus non-surgery and showed no change over 2 years(Kadanka, 2000). Their method of assessment was videod 10-m walk, mJOA, self assessment by patients. This study had ethical issues because all the patients had consented for surgery but only half received surgery. To be allowable, they would have selected only mildly affected and chronic patients (their earlier paper indicates that they were selected out of a group of 61 patients as being the ones less severely affected) (Bednarik and Kadanka, 1999), perhaps a group that many would not have considered operating upon. They had to have a JOA of atleast 12 and tended to have a very long duration of symptoms, up to about 30 years. Given the population group, it is perhaps not surprising that they revealed no benefit from surgery.

The previous chapters of this thesis have explored in detail the relative merits of various quantitative assessment measures for CSM, and described the possible advantages of a much more objective approach, namely actually quantifying walking performance. Such an approach is an extension of the early scales that focussed upon walking but merely graded it as "normal", "use of an aid" or "impossible". In addition to the previous chapter indicating the sensitivity of walking parameters to early operative changes (Singh and Crockard, 1999), a separate small study of gait analysis in 12 CSM patients showed an improvement in both walking speed and in step size following surgery (Kuhtz, Johnk et al., 1999).

This chapter outlines a study that hopes to give more definitive answers to questions on the role of surgical decompression in CSM management through a design that incorporates the following features:

1. Walking data are collected to allow an objective continuous range of assessment and more powerful analysis.
2. Most of the other scales are simultaneously employed for direct comparison with the maximum number of other studies / data sets.
3. The study is prospective and involves multiple follow-ups over defined periods similar in each patient, allowing serial measurements to look at early post-operative versus late post-operative (3 year) trends.

4. Comparison is made with a control population of patients with significant myelopathic severity who were not operated upon. An attempt is made to make a comparison of operated patients with *matched* unoperated patients.

6.3 Methods

The patients studied were consecutively referred for consideration of decompressive surgery for CSM to the Neurosurgical Department at the National Hospital for Neurology and Neurosurgery, Queen Square, UK. They had no condition other than CSM that would adversely affect their walking performance. A total of 60 patients were assessed pre-operatively but only 50 could be followed for the whole study duration. (The remainder are accounted for in the results.)

In addition, 34 "control" patients were recruited. These were all CSM patients who would have been offered decompressive surgery, but for reasons of personal preference or in consultation with their own doctors they chose not to have surgery at that time but nevertheless agreed to have their walking parameters assessed periodically. These patients were all aware that they could change their minds over surgery at any stage and indeed over the study duration 12 patients later elected for surgical decompression. This left 22 myelopathic "controls".

Local Ethical Committee approval was gained for this study and informed written consent was obtained from each patient and control after full explanation of the nature of the study. The protocol involved initial recording of the 30-m timed walk (time and number of steps taken) in the usual manner, and the recording at the same time of the JOA score, EMS, MDI score, Nurick score, Ranawat score and the SF36. Operations in the group who had elected for surgery were performed within a month of this assessment.

Repeat assessments were performed at 6 months, 1 year, 2 years and 3 years after the initial assessment. (The control patients completed the walking assessments but not the repeated battery of assessment scales.)

The data were analysed using appropriate statistical tests from the Microsoft Excel

and SPSS software analysis packages version 10.

6.4 Results

6.4.1 *Descriptive Results for Patients*

In all, 60 patients were assessed pre-operatively and followed up for three years after surgery. Of these, 50 patients were suitable for analysis of walking data. Of the remainder, 4 patients died of causes unrelated to CSM, one died immediately postoperatively, 2 were reoperated upon, one developed osteoporotic hip fractures that independently affected walking, one was wheelchair bound both pre- and post-operatively and one was wheelchair bound pre-operatively only.

The mean age (SD) of the 50 analysed patients was 56.7 (±13.68) years. There were 36 males and 14 females. A total of 29 patients received anterior decompressions while 7 had laminectomies and 14 laminoplasties.

As an initial rough indicator of the effects of surgery, only 1 patient out of 50 had worse walking time 6 months post-operatively while 2 were unchanged (i.e. same time to within 1 sec). On comparing 3 year follow up data with pre-operative times, 4 were worse and 3 were unchanged.

6.4.2 *Missing data:*

As described above, 10 patients could not be included in analysis for various reasons. These patients would not have importantly skewed the overall results. As an indicator of their performance, their walking times are illustrated (table 6.1).

Out of the remaining 50 patients, all data were available pre-operatively, 6 months post-operatively and upto 3 years post-operatively. One patient could not be measured at 1 year and 2 patients at 2 years. These patients were assigned values equidistant between the adjoining two assessment times.

Patient	Pre-op (s)	6 month post-op (s)	1 year post-op (s)	2 years post-op(s)	3 years post-op (s)
1	188	30	30	52 (fracture hip)	38
2	29	26	27	24	22 (re-op pain)[1]
3	126	96	166	108	118 (re-op walking worse)[2]
4	240	190	196	died of stroke	
5	38	34	66	Died of brain haemorrhage	
6	150	peripheral neuropathy	died IHD		
7	25	died of heart failure			
8	Wheelchair	died peri-op			
9	Wheelchair	10-m walking	wheelchair	wheelchair	5-m walking
10	Wheelchair	155	20-m walking	113	116

Table 6.1 Details on 10 patients that were not included in the data analysis, reasons for exclusion and their pre and post-op walking times.

[1] This patient's main symptom was arm and leg pain and she was successfully re-operated upon for recurrence.

[2] MRI revealed that decompression was incompletely relieved following initial surgery. His walking deteriorated dramatically after the 2-year assessment and improved again following a second decompressive operation.

6.4.3 Analysis of Walking times

The mean (95% confidence interval) walking time pre-operatively was 53.6 (10.3) s, that at 6 months post-operatively was 39.9 (6.6) s, that at 1 year was 41.2 (7.2) s, that at 2 years was 40.9 (7.4) s, and that at 3 years was 38.6 (6.9) s. All the four post-operative assessment occasions thus showed faster mean walking times (fig. 6.1). Repeated measures analysis of variance (SPSS) was performed upon the fifty patients' five sets of walking times. Sphericity could not be assumed and so a Greenhouse-Geisser correction was performed (reducing the degrees of freedom from 4 to 1.6). This gave a more conservative statistical measure. A highly significant difference in walking times across the five assessment occasions was found ($p<0.001$). On making individual contrasts against pre-operative times, all 4 sets of post-operative walking times were improved (six months $p<0.001$; one year $p=0.001$; two years $p<0.001$; three years $p<0.001$). Repeated contrasts, looking at changes between consecutive assessment occasions, revealed as expected a significant improvement from pre-op to 6 months post-op ($p<0.001$), no change from 6 months to 1 year ($p=0.306$), no change from 1 year to 2 years ($p=0.683$) and a suggestion of a late *improvement* from 2 years to 3 years ($p=0.02$). (This last value is considered non-significant because of the multiple comparisons made.)

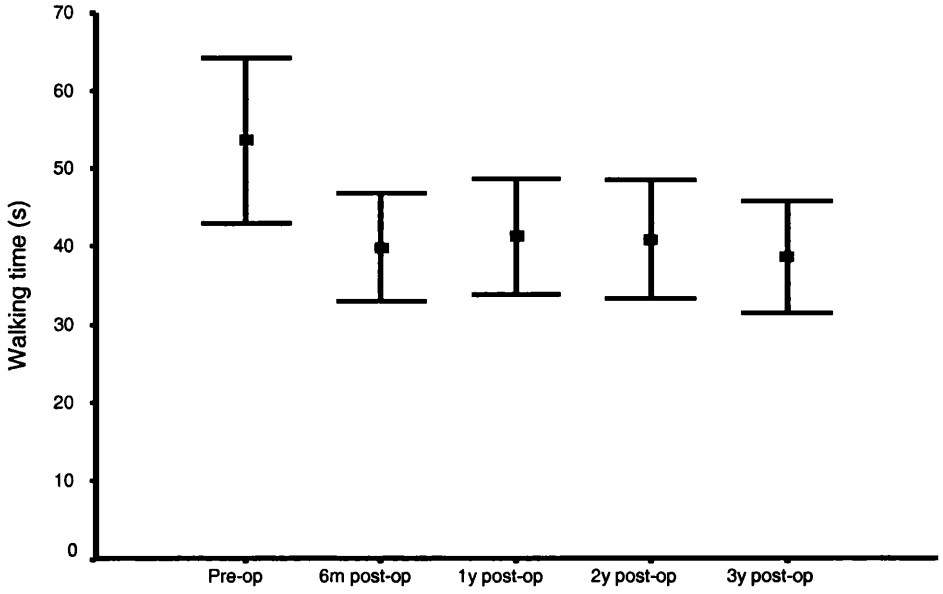


Figure 6.1 Means of walking times for 50 patients pre-operatively and for 4 follow-up occasions post-operatively. The improvement in walking times following surgery on all four occasions was statistically significant.

An important aspect of long-term follow up is the possibility that patients once operated upon may deteriorate again over time to their pre-operative level. This is not suggested by the contrast analysis above and is formally demonstrated by a separate test showing lack of significant difference between the four post-operative values ($p=0.163$) and the lack of any linear trend across these values ($p=0.404$).

6.4.4 Analysis of number of steps taken

The mean (95% confidence interval) number of steps taken pre-operatively was 64.8 (6.0), that at 6 months post-operatively was 56.4 (4.6), that at 1 year was 56.4 (5.1), that at 2 years was 56.6 (5.4), and that at 3 years was 55.5 (5.8). Thus, all the four post-operative assessment occasions showed fewer steps taken than pre-operatively (fig 6.2). Repeated measures analysis of variance (SPSS) was performed upon the fifty patients' five sets of steps taken. Again, a highly significant difference in steps taken across the five assessment occasions was found ($p<0.001$, with Greenhouse-Geisser correction reducing the degrees of freedom from 4 to 2.6).

On making individual contrasts against pre-operative steps, all 4 sets of post-operative steps taken were improved (six months $p<0.001$; one year $p<0.001$; two years $p<0.001$; three years $p<0.001$).

Repeated contrasts, looking at changes between consecutive assessment occasions, revealed as expected a significant improvement from pre-op to 6 months post-op ($p<0.001$), no change from 6 months to 1 year ($p=0.985$), no change from 1 year to 2 years ($p=0.82$) and no change from 2 years to 3 years ($p=0.273$). There was no significant difference between the four post-operative values ($p=0.728$) and a lack of any linear trend across these values ($p=0.629$).

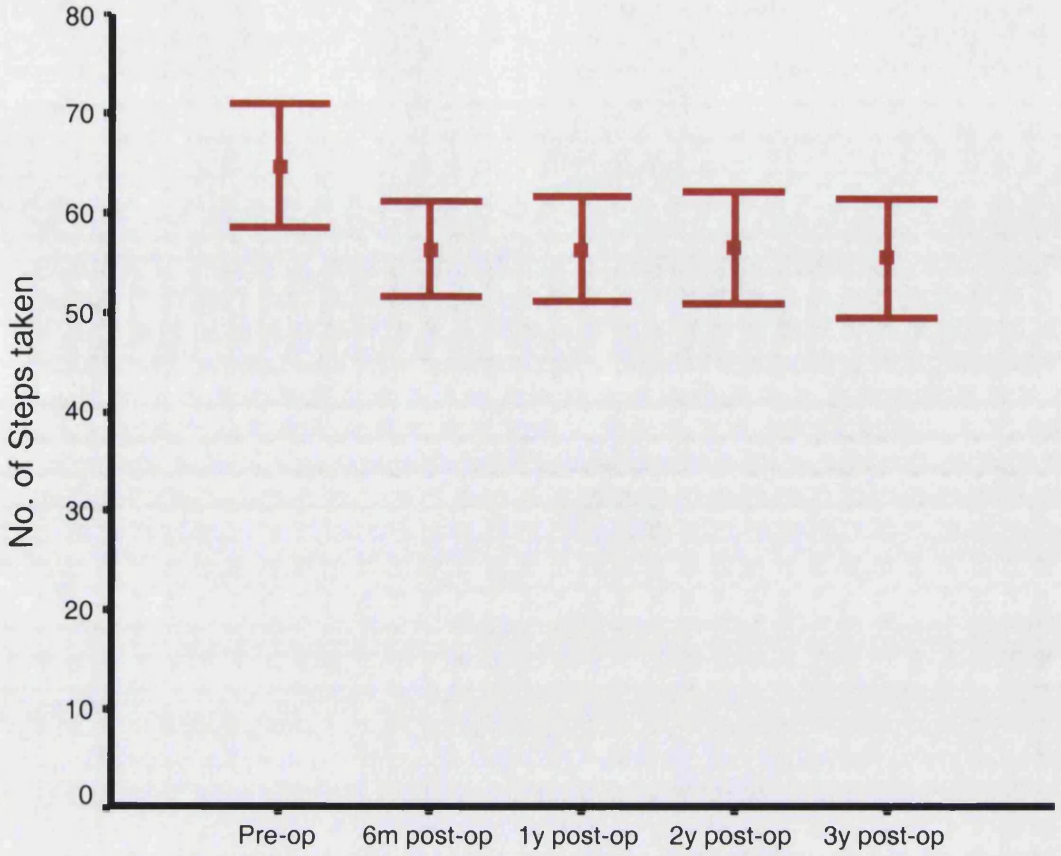


Figure 6.2. Means of number of steps taken to walk 30-m for 50 patients pre-operatively and for 4 follow-up occasions post-operatively. The improvement in number of steps taken to walk 30-m following surgery on all four occasions was also statistically significant.

6.4.5 ***Comparison with unoperated myelopathic controls:***

To determine the true effects of surgical intervention, a control group of patients with CSM was followed up over a similar three year period. For ethical reasons these patients of course had access to neurosurgical intervention if they changed their minds at any stage and elected to have surgery.

In the event, out of 34 patients in the control group, 12 patients over the three years had surgery. This was because they and their doctors felt that their condition had deteriorated to an unacceptable level. The remaining 22 patients were analysed. This clearly introduced a bias where the control group would appear to have a more favourable outcome.

Nevertheless these 22 patients still deteriorated in walking time over the 3 years.

At 6 months, 9 patients deteriorated while 5 remained unchanged and 8 showed a modest improvement. By 3 years, 19 had deteriorated while only three had improved. Of the three who improved, one said it was the result of daily massage to his neck and the wearing of a soft collar, one said he had adapted to his problem and the third said he was having regular manipulation treatment and yoga.

The age (95% confidence interval) of the control patients was 53.0 (5.1) years. The mean (95% confidence interval) walking time initially was 38.8 (7.2) s, at 6 months was 40.7 (8.1) s, at 1 year was 43.2 (7.7) s, at 2 years was 45.0 (9.2) s and at 3 years was 51.8 (11.0) s (fig. 6.3). Repeated measures analysis of variance (SPSS) using the Greenhouse-Geisser correction gave a significant deterioration over time (p=0.002). Taking individual contrasts over time revealed that the 3 year assessment occasion had significant deterioration compared with the initial occasion (6 months p=0.19; 1 year p=0.027; 2 years p=0.045; 3 years p=0.003). Contrasts comparing consecutive occasions revealed a significant deterioration between 2 years and 3 years (6 months to 1 year p=0.078; 1 year to 2 years p=0.23; 2 years to 3 years p=0.001). Finally, there was a significant linear trend of deterioration over time (p=0.005).

For a direct comparison of control patients with decompressed patients, the 22 control patients were severity matched on the basis of initial walking time, sex matched and, as far as possible, age matched with 22 of the 50 operated patients. A good match was achieved by blinded observation of the data. The operated subgroup had a mean (95% confidence interval) age of 53.2 (4.0) years and a preoperative walking time of 39.2 (7.1) s. These are very close to the control values (53.0 (5.1) years and 38.8 (7.2) s).

The walking times of this subgroup of decompressed patients changed similarly to the *overall* group of decompressed patients and over the assessment occasions became clearly better than the matched control patients (fig. 6.3). Simple comparison of means (unpaired t-test, SPSS) showed possible differences at 6 months, 1 and 2 years (given correction for multiple comparisons) and a clear difference at 3 years (initial assessment p=0.93; 6 months p=0.07; 1 year p=0.02; 2 years p=0.02; 3 years p=0.003).

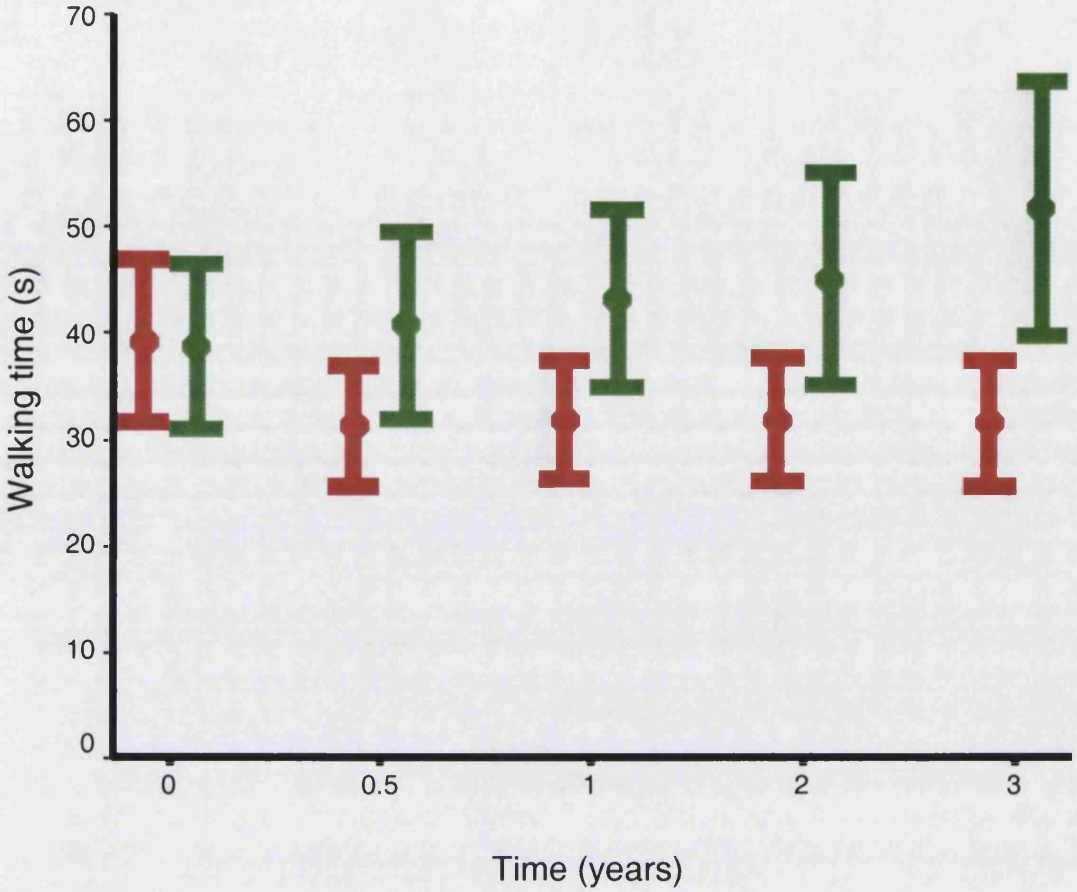


Figure 6.3. Walking times (with confidence intervals) for 22 severity, sex and age matched decompressed patients (red bars) with unoperated control patients (green bars). The initial walking times were similar, reflecting close matching, but the operated patients' walking times after surgery became progressively better than the control group of patients.

A more sensitive comparison is to use the matched pairs of untreated versus decompressed patients. Thus, the *differences* in walking times between each pair of patients were compared over the five occasions. The means of the differences changed significantly over the 5 assessment occasions (analysis of variance with Greenhouse-Geisser correction (df=2.06), $p<0.001$). On making contrasts over consecutive assessment occasions, there was a significant deterioration of controls versus operated patients (repeated measures) at 6 months and a further significant widening of this deterioration at 3 years (initial to 6 months $p<0.001$; 6 months to 1 year $p=0.15$; 1 year to 2 years $p=0.26$ and 2 years to 3 years $p<0.001$). There was a significant linear trend of increasing

separation in walking time between decompressed and untreated patients ($p<0.001$). Comparing the mean differences with zero difference rather than with each other revealed significant results after the initial assessment (initially $p=0.26$ (where the patients were non-significantly worse due to imperfect matching); 6 months $p<0.001$; 1 year $p<0.001$; 2 years $p<0.001$; 3 years $p<0.001$). Thus, surgery improved patients' walking times and these improvements were not only maintained but actually increased over the three years following surgery.

6.4.6 *Effect of Age upon Walking Time:*

The preliminary study of walking data (chapter 4) indicated that there may be an effect of age on walking time where older subjects walk slower. To identify the contribution of *normal* ageing on walking time, the initial study assessed walking time in 40 normal controls of ages similar to a CSM population. By linear regression, it was found that older patients indeed walked more slowly at the rate of 0.12 seconds/ year. An appropriate reference point for age was 60 years (a median value). Thus, one can correct for age by subtracting 0.12 s/ year for patients over 60 and adding a similar amount for patients under 60.

Since the match between unoperated patients and decompressed patients was not perfect for age, this correction was applied, where t is the walking time:

$t_c = t + 0.12(60 - \text{initial age} - \text{year into study})$

Matching of the age-corrected untreated patients and the age-corrected decompressed patients resulted in two changes to the 22 pairs of matches (i.e. selecting a different patient out of the 50 possible) and one swap. Comparison of means (unpaired t-test, SPSS) showed possible differences at 6 months, 1 and 2 years (given correction for multiple comparisons) and a clear difference at 3 years (initial assessment $p=0.90$; 6 months $p=0.038$; 1 year $p=0.012$; 2 years $p=0.015$; 3 years $p=0.002$). The magnitudes of the differences (fig. 6.4) became if anything greater than for uncorrected walking times.

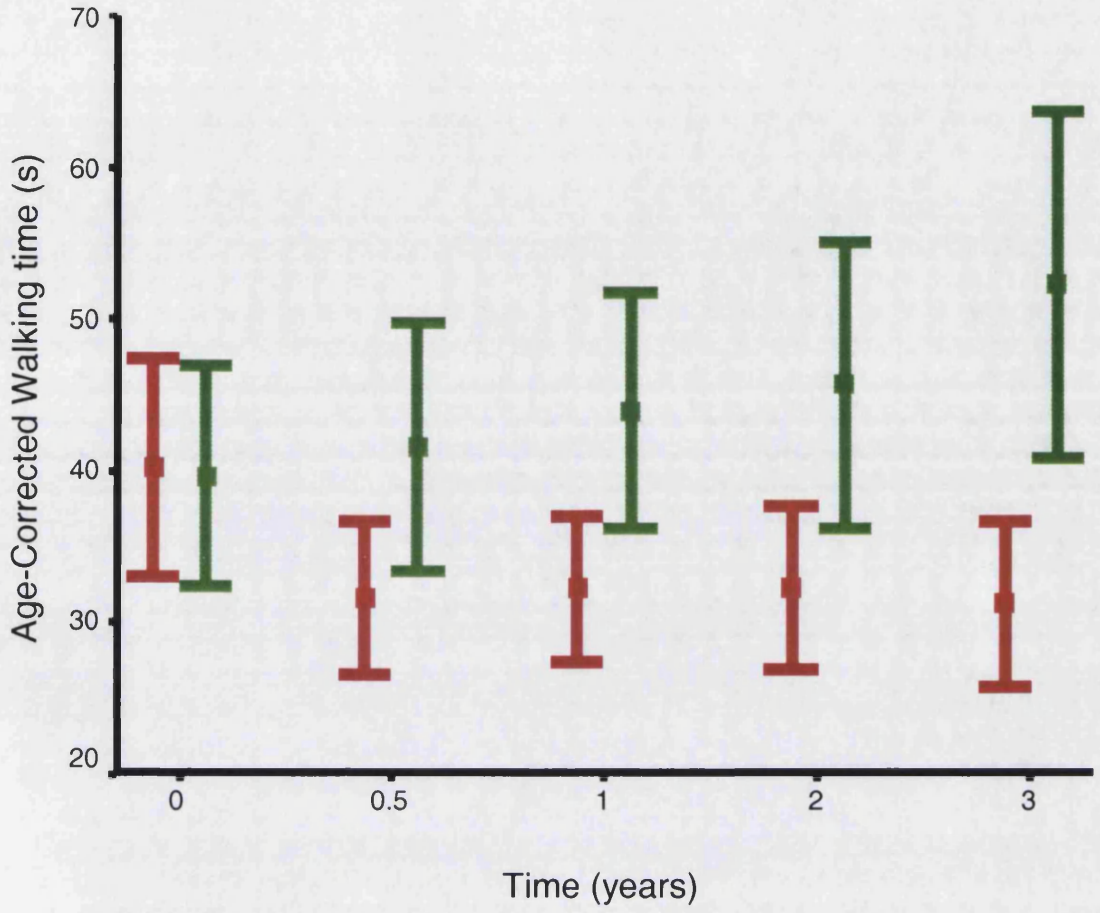


Figure 6.4. Mean differences (with confidence intervals) in the walking times of the severity and sex matched 22 decompressed patients (red bars) with unoperated controls (green bars) after age correction. There are significant differences over the 4 post-op occasions.

The means of the differences between each pair of patients changed significantly over the 5 assessment occasions (analysis of variance with Greenhouse-Geisser correction (df=2.28), $p<0.001$). On making contrasts over consecutive assessment occasions, there was again a significant deterioration of controls versus operated patients at 6 months and a further significant widening of this deterioration at 3 years (initial to 6 months $p<0.001$; 6 months to 1 year $p=0.27$; 1 year to 2 years $p=0.28$ and 2 years to 3 years $p<0.001$). There was a significant linear trend of increasing separation in corrected walking time between decompressed and untreated patients ($p<0.001$). Comparing the mean differences with zero difference again revealed significant results after the initial assessment (initially

p=0.26; 6 months p<0.001; 1 year p<0.001; 2 years p<0.001; 3 years p<0.001).

6.4.7 ***Comparison of number of steps taken with unoperated myelopathic controls:***

The number of steps taken by the same unoperated control patients over the 30-m course were similarly compared with a matched subgroup of decompressed patients. The patients giving the best match for initial number of steps taken were somewhat different from those best matching for initial walking time.

The data for unoperated controls (fig. 6.5) was similar to that for walking times, although there appeared to be a slight and non-significant *improvement* in steps taken at 6 months, which later deteriorated again.

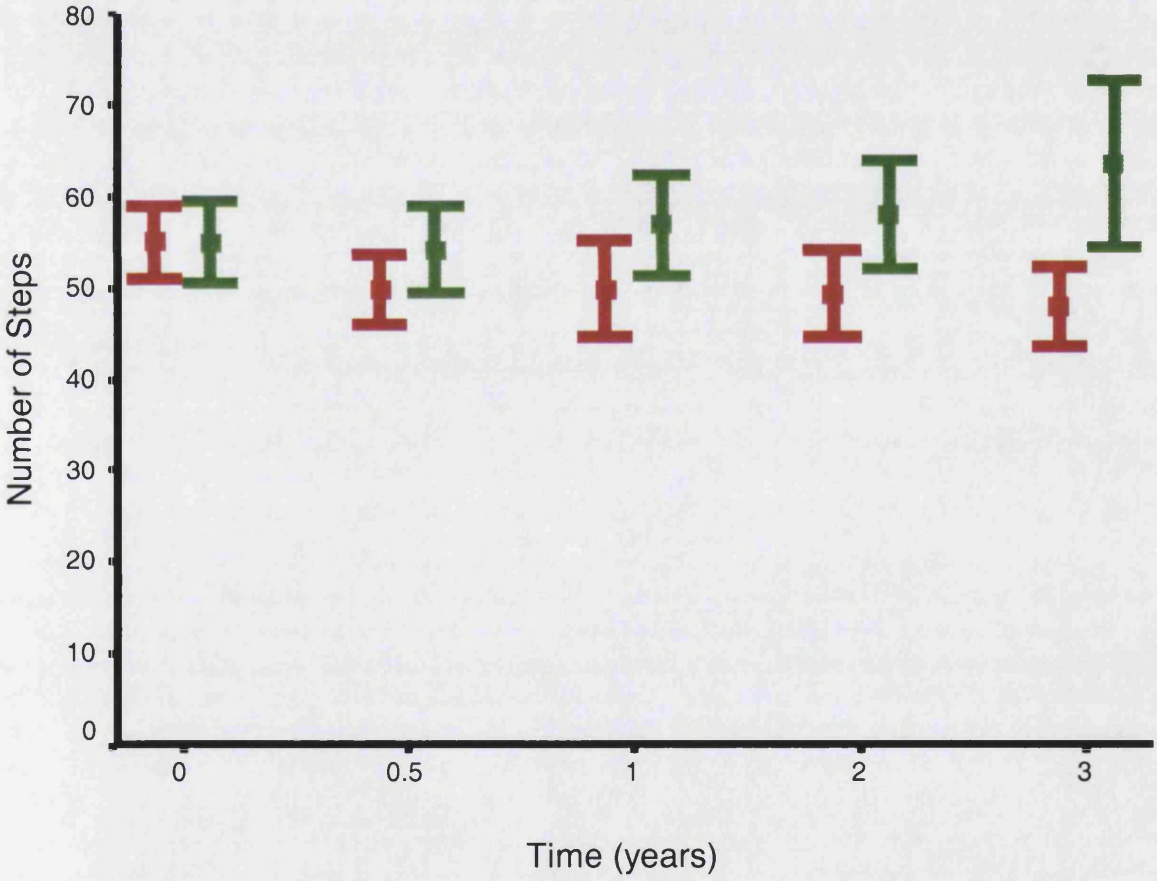


Figure 6.5 Number of steps taken (with confidence intervals) to walk 30-m for the severity, age and sex matched 22 decompressed patients (red bars) with unoperated control patients (green bars). When comparing the difference in steps taken for the two groups over the 4 post-op occasions, there was a significant linear trend showing increasing difference in the steps taken between the two groups.

Comparison of means (unpaired t-test, SPSS) showed a possible significant difference at 2 years (given correction for multiple comparisons) and a clear difference at 3 years (initial assessment p=0.99; 6 months p=0.14; 1 year p=0.065; 2 years p=0.025; 3 years p=0.004).

The means of the differences in steps taken between each pair of patients changed significantly over the 5 assessment occasions (analysis of variance with Greenhouse-Geisser correction (df=2.06), p=0.002). On making contrasts over consecutive assessment occasions, there was less clear progressive deterioration of controls versus operated patients than for walking times (initial to 6 months p=0.011; 6 months to 1 year p=0.19; 1 year to 2 years p=0.48 and 2 years to 3 years p=0.025). There was still a significant linear trend of increasing separation of steps taken between decompressed and untreated patients (p=0.004). Comparing the mean differences with zero difference revealed rather less significant results after the initial assessment than for walking time (initially p=0.93; 6 months p=0.013; 1 year p=0.024; 2 years p=0.01; 3 years p=0.003).

Thus the number of steps taken revealed the same significant deterioration of unoperated versus decompressed patients albeit rather less strongly than did walking times. Using age correction for number of steps taken (0.21 steps/year) did not materially affect these results (not illustrated).

6.4.8 Severity effect on Improvement

An important question is whether patients do better from surgery when they are more mildly affected or do better when operated upon with severe disease. This was addressed by linear regression analysis (Fig. 6.6) of walking time improvement on initial pre-operative walking time. There was a highly significant effect where patients with *slower* walking times had *greater* improvement (Confidence Interval) from surgery at 6 months (slope =0.50 (0.38-0.62), $p=4.67 \times 10^{-11}$). This meant that for every 10 seconds slower a patient was before surgery, he improved on average an extra 5 seconds afterwards. This effect was maintained at 3 years post-operatively (slope = 0.47 (0.35-0.58), $p=2.4 \times 10^{-10}$). Analysis of age-corrected data did not affect these findings (6 months

slope = 0.51(0.39-0.63), $p=2.2x10^{-11}$); 3 year slope = 0.48 (0.36-0.59), $p=1.7x10^{-10}$).

Taking *relative* improvement (i.e. (pre-op time minus post-op time)/ pre-op time) rather than *absolute* improvement yielded the same effect but was less clear (6 months p=0.01; 3 years p=0.03).

Finally, the influence of initial severity on the *unoperated* patients' deterioration was explored. Worse initial severity tended to lead to worse deterioration from initial walking time to 3 year walking time, but this was non-significant (p=0.53).

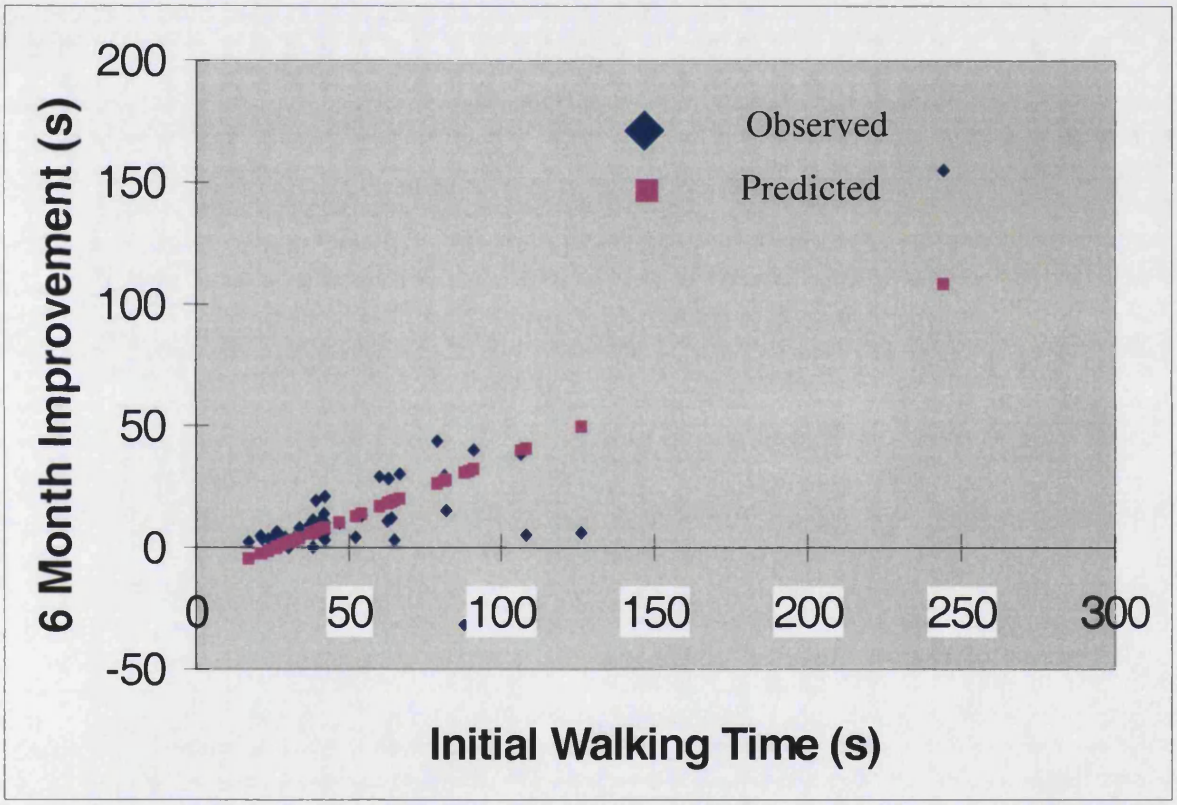


Figure 6.6 Linear regression analysis of walking time improvement at 6 months on pre-operative walking time.

6.4.9 Age effect on improvement

Increasing age might lead some surgeons to expect a poorer result from surgery. However, analysis of these data revealed this was not the case. Linear regression of absolute improvement on age revealed a tendency for *greater* benefit in older patients,

although this was not significant (6 months p=0.58; 3 years p=0.26). Obviously similar non-significant trends occurred for relative improvement.

In contrast, on looking at the unoperated patients, there was worse deterioration from initial walking time to 3 year walking time in older patients (slope = -0.67 (-0.08 to –1.23) s/year, p=0.028). A patient 10 years older was likely average to deteriorate by an extra 6.7 s in walking time over 3 years.

Thus older CSM patients did at least as well from surgery and deteriorated faster if not operated upon.

(There may be a potential interaction between the age and severity effects, since older patients may initially be more severely affected (see chapter 4). However, the effect of initial severity on improvement was unchanged when corrected for age (previous page). The findings for age and severity were different in any case – worse patients significantly improved from surgery but worse controls had unchanged deterioration while older controls deteriorated more but older operated patients did not have significantly greater surgical benefit.)

6.4.10 Effect of Surgical Procedure on Improvement

Three main types of surgical procedure were performed upon the operated patients in this study. The absolute improvement in walking times for the anterior decompressions (29 patients), posterior laminectomies (7 patients) and posterior laminoplasties (14 patients) were compared. At 6 months, the mean improvement for the three procedures was 11.0 (4.2) s, 27.3 (55.4) s and 12.6 (6.8) s respectively. Although there was a suggestion that posterior laminectomy was better than the other two procedures, the wide variation and small number involved meant there was no significant difference or contrasts between these means (p=0.27, one way analysis of variance independent samples). At three years, the mean improvements were 10.4 (5.0) s, 26.8 (44.8) s and 15.0 (6.5) s. There was again no significant difference or contrasts (p=0.19).

6.4.11 Other Assessment Measures

To corroborate with the walking data, the decompressed patients were also scored on a number of scales that have been used to quantify myelopathic severity.

6.4.11.1 Japanese Orthopaedic Association score

The JOA disability scale ranges from 0 to 17 (normal). The median (quartile) score preoperatively was 9 (8, 12) and the postoperative scores at 6 months, 1 year, 2 years and 3 years were 10 (8.75, 14), 11 (9, 15), 11 (9, 13.25) and 10 (8.75, 14) respectively; these all indicated a significant improvement (Wilcoxon signed rank test, SPSS) over pre-operative levels ($p=0.001$; $p<0.001$; $p<0.001$; $p<0.001$ respectively). Overall improvement across the 5 assessment occasions was also significant ($p<0.001$, Friedman, SPSS). The worsening at 3 years compared to 2 years was not significant ($p=0.37$, Wilcoxon signed rank test). A "recovery rate %" is sometimes calculated from the relative change in JOA score (Hirabayashi, Miyakawa et al., 1981). This was done at 6 months (fig. 6.7) and at 3 years (fig. 6.8)

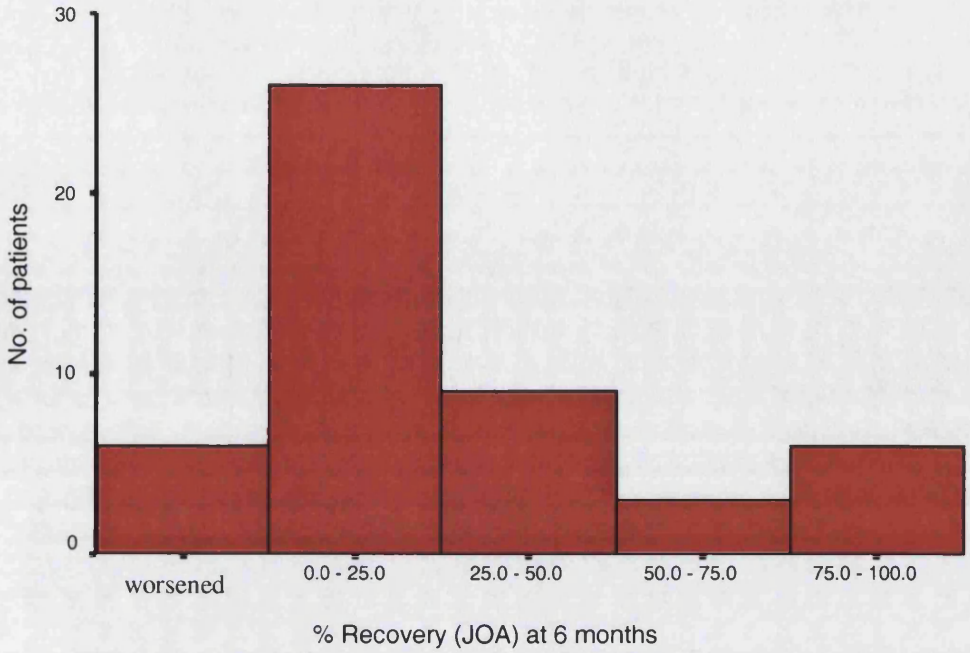


Figure 6.7 Histogram for '% recovery rate of JOA' for 50 decompressed patients at 6 months following surgery. Most patients showed 'poor improvement' i.e. 0-25% at 6 months.

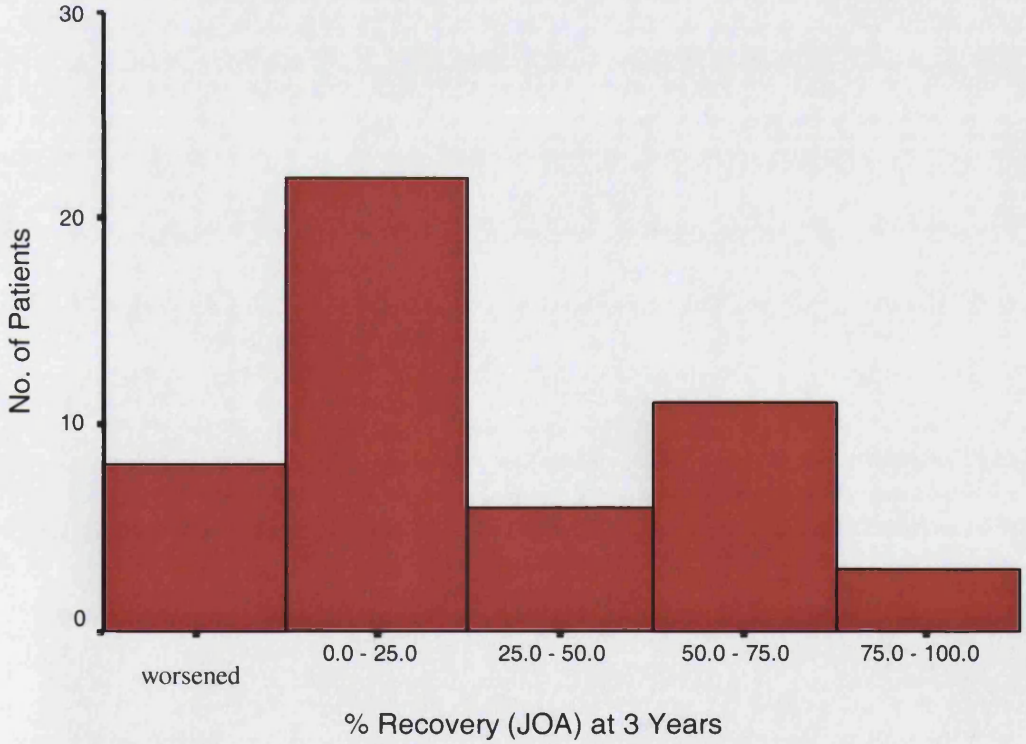


Figure 6.8 The histogram for '% recovery rate of JOA' for 50 decompressed patients at 3 years after surgery. Most patients showed 'poor improvement' i.e. 0-25%, a similar finding to that at 6 months.

This recovery rate was used to compare the relative effects of the three main type

of surgery performed: anterior decompression, posterior laminectomy and posterior laminoplasty. At 6 months the mean recovery % for the three procedures was 25.6 (13) %, 27.5 (28.8) % and –3.75 (32.7) %. When contrasted against the other two procedures, the posterior laminoplasties did worse (p=0.036). Some of this effect may have been the large bias for negative (worse) results that occurs when calculating recovery %. When the % values were converted into excellent, good, fair, poor and worse grades, the median results for the three procedures (quartiles) were poor (poor, fair), fair (poor, fair) and poor (poor, poor). The non-parametric Kruskall-Wallace test did not reach significance comparing the three (p=0.1). Note that this pattern of results was different from the walking time results on the same patients.

At three years, there were no significant differences between operations, but a trend for anterior decompression to be best, followed by laminectomy and then laminoplasty. (Means were 23.3 (19) %, 14.1 (33.3) % and 11.3 (20.6) % (p=0.68, no contrasts). For grades, medians were fair (poor, good), poor (poor, poor) and poor (poor, fair) (p=0.30).

6.4.11.2 European Myelopathy Score

This disability scale adapted from the JOA ranges from 5 (worst) to 18. Point totals 17 and 18 are considered normal. The median (quartile) score preoperatively was 12 (10, 14) and the postoperative scores at 6 months, 1 year, 2 years and 3 years were 14.5 (12, 17), 14 (12, 17), 13.5 (12, 16) and 14 (12, 17) respectively; these all indicated a significant improvement (Wilcoxon, SPSS) over pre-operative levels ($p<0.001$ in all four cases). Overall improvement across the 5 assessment occasions was also significant ($p<0.001$, Friedman, SPSS).

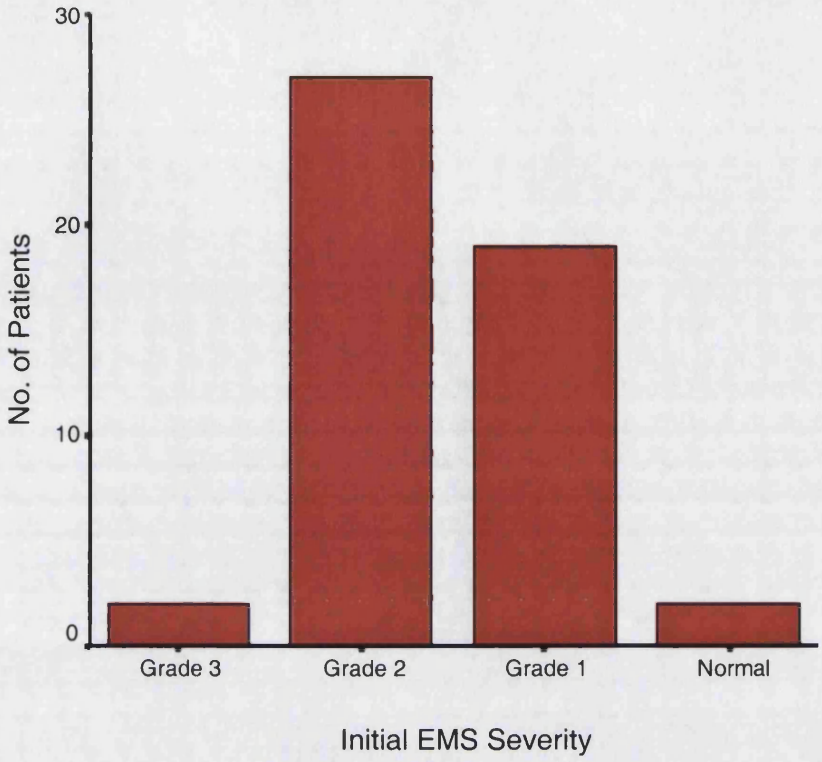


Fig. 6.9 Histogram of initial EMS severity grades.

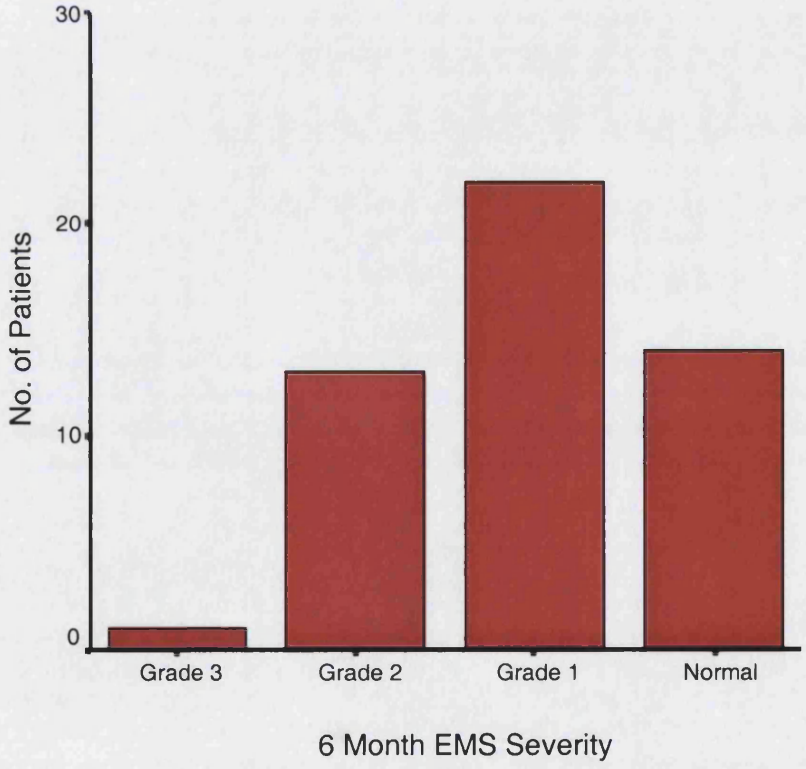


Figure 6.10 Histogram of EMS severity grades at 6 months.

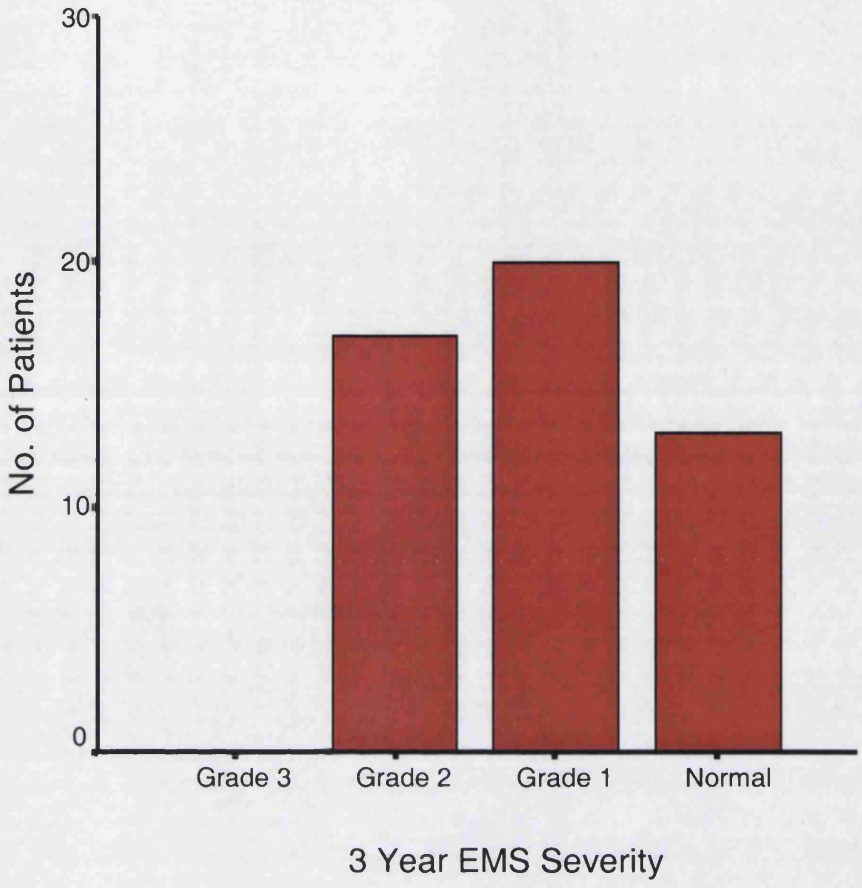


Figure 6.11 Histogram of EMS severity grades at 3 years

The EMS is typically divided into grade 3 (5-8 points), grade 2 (9-12 points), grade 1 (13-16 points) and normal (17-18 points). Initial, 6 month and 3 year EMS severity grades were determined (figs. 6.9, 6.10, 6.11). A greater proportion of patients were graded normal rather than severe after the surgery, illustrating the maintained improvement over time.

6.4.11.3 Myelopathy Disability Index

The MDI ranges from 0 (no disability) to 33 (worst disability) and is generally converted into a percentage. The median (quartile) score preoperatively was 35 (17, 48) % and the postoperative scores at 6 months, 1 year, 2 years and 3 years were 14 (6, 34) %, 17 (6, 34) %, 18 (9, 31) % and 12 (3, 30) % respectively (fig. 6.12). These all

indicated a significant improvement (Wilcoxon, SPSS) over pre-operative levels ($p<0.001$ in all four cases). Overall improvement across the 5 assessment occasions was also significant ($p<0.001$, Friedman, SPSS). However, there was no significant difference between the 4 post-operative assessment occasions.

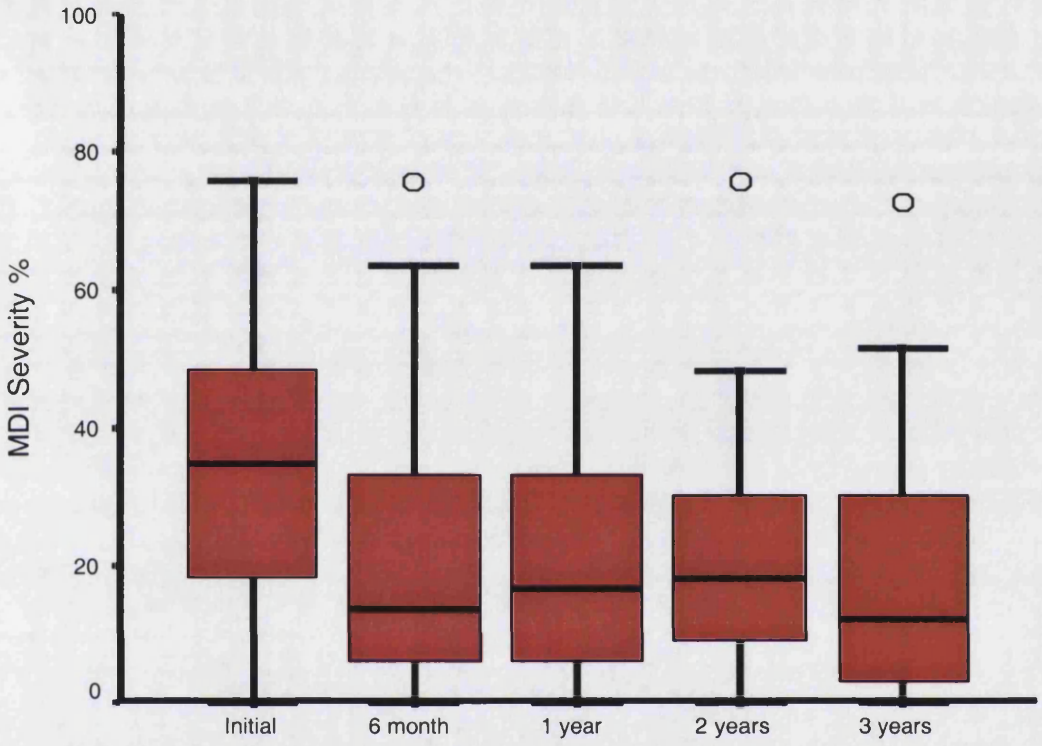


Figure 6.12 Box plots of MDI scores for 50 decompressed patients pre-operatively and post-operatively. The circles represent outlying values greater than 1 ½ interquartile intervals. In all cases, the improvement following surgery was statistically significant (Wilcoxon SPSS $p<0.001$).

6.4.11.4 Nurick Functional Scale

This scale is an impairment and functional measure ranging from grade 0 (no signs of spinal cord disease) to 5 (chair bound). The median (quartile) score preoperatively was 3 (3, 4) and the postoperative scores at 6 months, 1 year, 2 years and 3 years were 2 (1, 2), 2 (1, 3), 2 (1, 3) and 2 (1, 3.25) respectively. These all indicated a significant improvement (Wilcoxon, SPSS) over pre-operative levels ($p<0.001$ in all four cases). There was however a strong suggestion that function deteriorated again from 6 months to 3 years ($p=0.015$). This is in contradiction with the other scales, particularly the walking

tests, and is surprising in view of the fact that grades 2 to 5 are largely concerned with walking ability. Perhaps over time, patients re-interpreted walking difficulty sufficient to "prevent full-time employment or housework" or adopted the use of a walking stick (grade 4) which may actually have improved their walking performance.

6.4.11.5 Ranawat Scale

This is another combination impairment / walking scale graded I (normal) to IV (chair bound). The median (quartile) score preoperatively was 3 (3, 3) and the postoperative scores at 6 months, 1 year, 2 years and 3 years were 2 (1, 2), 2 (2, 3), 2 (2, 3) and 2 (2, 2) respectively. These all indicated a significant improvement (Wilcoxon, SPSS) over pre-operative levels ($p<0.001$ in all four cases). Again, there was a significant deterioration again from 6 months to 3 years ($p=0.001$). (The worst two grades on the Nurick and Ranawat scales constitute a near identical criteria.)

6.4.11.6 Short Form 36 Questionnaire

The SF36 questionnaire, a measure of subjective disability and handicap, was recorded for each patient on all five assessment occasions. The 36 questions are analysed as eight transformed aspects of function (interval measures defined as % of normal) and one ordinal function of health change (grade 1 to 5).

The health change function had a median (quartiles) preoperatively of 4 (3, 4) and improved post-operatively at 6 months, 1 year, 2 years and 3 years to 2 (1, 3), 2 (1, 3), 3 (2, 3) and 3 (2, 3) respectively. All these values were significantly better than preoperative scores ($p<0.001$, Wilcoxon, SPSS). The late deterioration from 6 months to 2 and 3 years was of borderline significance ($p=0.039$, Friedman, SPSS).

The means for the other eight categories all improved at 6 months but tended to deteriorate again later to varying degrees (fig.6.13). The significance of these changes is shown in table (6.2). (Note that multiple comparison corrections have not been made.)

Measure	Improvement pre-op to 6 month	Improvement pre-op to 3 year	Deterioration (linear trend) from 6 month to 3 year
General Health	P=0.12	P=0.72	P=0.007
Physical Function	P<0.001	P<0.001	P=0.57
Role Physical	P=0.001	P<0.001	P=0.33
Role Emotional	P<0.001	P<0.001	P=0.82
Social Function	P<0.001	P=0.006	P=0.068
Bodily Pain	P<0.001	P=0.159	P=0.046
Mental Health	P<0.001	P=0.001	P=0.26
Vitality	P<0.001	P=0.001	P=0.018

Table 6.2. The post-operative improvement and deterioration in the 8 categories of SF36.

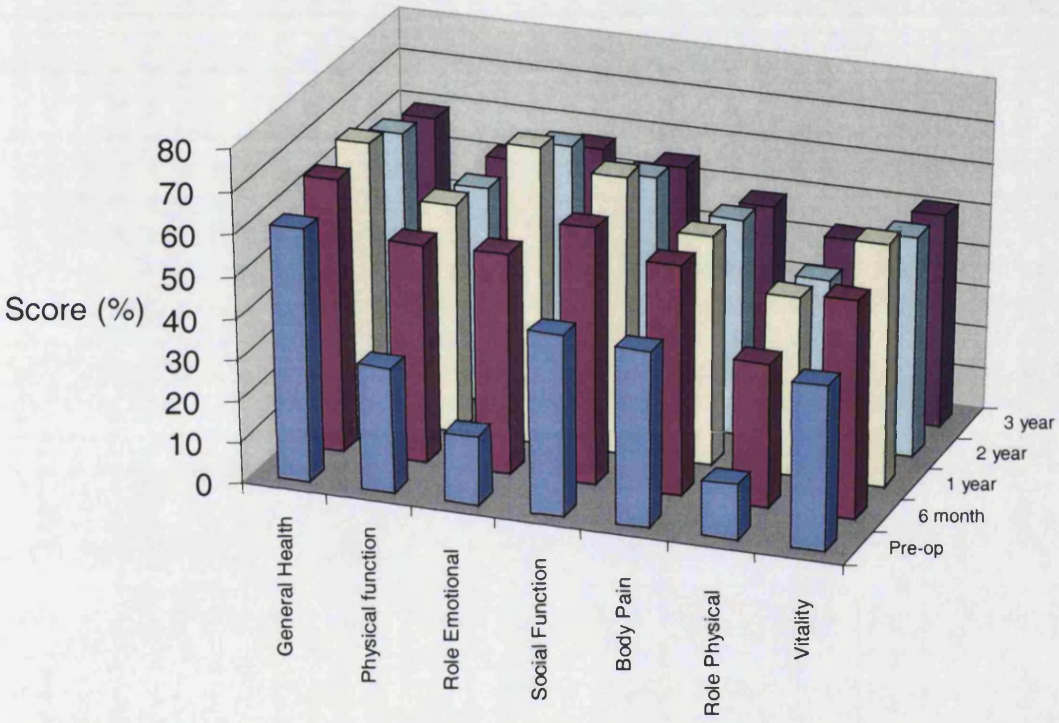


Figure 6.13 Scores (% of normal) for 8categories of SF36 pre-operatively and post-operatively at 4 follow-up sessions. Generally there is improvement at 6 months and then some deterioration by 3 years, but not as bad as the pre-operative level.

There was no significant early or late improvement in general health and a trend for its deterioration after its non-significant improvement following surgery. Body pain early on improved following decompression but deteriorated so that it was not significantly better than preoperative levels at 3 years. The remaining aspects all showed sustained improvement, although vitality and possibly social function may have had a significant trend for deterioration over the 6 month to 3 year period, but nevertheless still

better than preoperative levels.

6.4.12 Bladder function

Bladder dysfunction is considered an important criterion of myelopathic disease that may warrant early surgical decompression. The problems with the assessment of bladder function by means of scales have already been outlined in the previous chapter where it was found that scale sub-components looking specifically at bladder function correlated poorly, presumably because of the wording of the question (Singh and Crockard, 2001).

Nevertheless, to investigate if bladder function behaved significantly differently from walking parameters and the scales, the bladder components of the JOA score were analysed separately. This score ranges from 0 (worst) to 3 (normal). It was found that most patients scored normally both pre- and post-operatively. Thus the medians were all normal and there was no significant difference between the 5 occasions of assessment (Friedman, p=0.39). On making individual comparisons between pre-operative and 3 year post-operative values, 37 patients were normal throughout, 6 were abnormal and remained at the same severity, 5 deteriorated and 2 improved. There were therefore too few number to make judgements, although there was an indication that surgery did not yield the same kind of benefits for bladder function as for walking and overall function.

One must note that the case mix in this study largely constituted patients with chronic myelopathy. In such patients, severe bladder dysfunction appears not to be a common problem. Of course, in this age group it is difficult in any case to distinguish myelopathic bladder dysfunction from prostatism or gynaecological bladder problems. Bladder dysfunction may be more of a warning feature relating to myelopathy resulting from *acute* cord compression. However, this thesis has deliberately not focussed on such patients since the decision to operate and the timing of surgery is relatively obvious compared to the *chronic* CSM patient.

6.4.13 Neck and Arm Pain

A further aspect of cervical spondylosis that this thesis has not focussed upon is pain, largely because this is mainly a local (neck) or radiculopathic (arm) symptom rather than myelopathic feature. However, such symptoms may still be important for individual patients and it would be important to know if surgery had adverse effects on such symptoms. While some overall scales include pain questions (JOA scale, EMS and SF36) and these showed both overall improvement and pain subsection improvement (SF36 figure 6.13 in this chapter and JOA & EMS previous chapter), the walking data clearly do not. Therefore a separate pain analogue score test (0 (no pain) to 10 (most severe imaginable) was performed on each patient on each of the 5 occasions (figs, 6.14, 6.15).

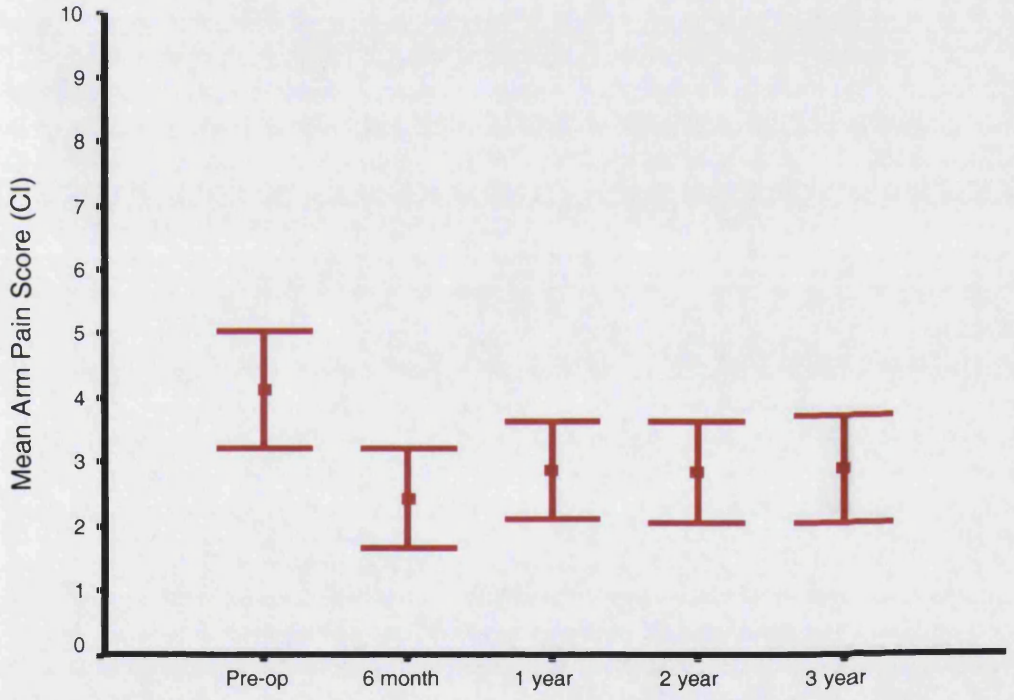


Figure 6.14 Pre-operative and post-operative neck pain (visual analogue score). The improvement following surgery was similar to the walking data.

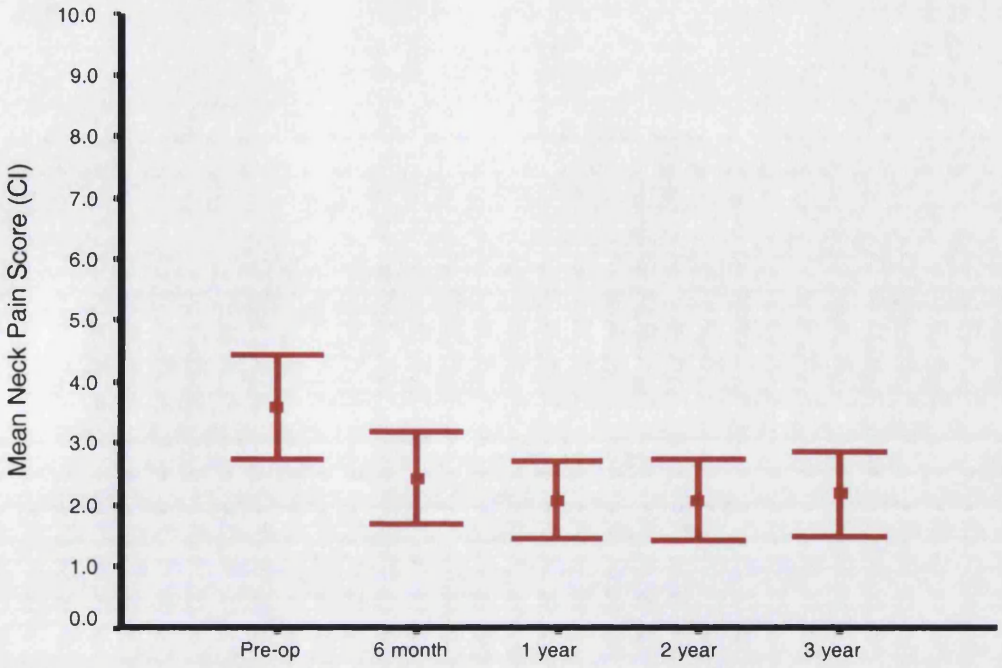


Figure 6.15 Pre-operative and post-operative arm pain (visual analogue score). The improvement following surgery was similar to the walking data.

Both neck pain and arm pain showed improvement from surgery in a pattern similar to that for walking data (means for neck pain pre-op and at 6 months, 1 year, 2 years and 3 years: 3.6, 2.4, 2.1, 2.1, 2.2; means for arm pain: 4.1, 2.4, 2.9, 2.8, 2.9). (Confidence intervals are illustrated in the figures.)

Regarding neck pain, there was a significant overall change ($p<0.001$ with Greenhouse Geisser correction). Contrasts showed that this was in each case between the pre-operative and four post-operative values ($p=0.011$, $p=0.001$, $p<0.001$, $p=0.001$). There was no suggestion of a trend of late post-operative improvement or deterioration from 6 months to 3 years.

Regarding arm pain, there was again a significant overall change ($p=0.002$ with Greenhouse Geisser correction). Contrasts showed that this was in each case between the pre-operative and four post-operative values ($p<0.001$, $p=0.013$, $p=0.003$, $p=0.012$). There was a suggestion of a trend of late post-operative deterioration from 6 months to 3 years, but this was not significant (6 month, 1 year, 2 year and 3 year comparisons, $p=0.475$; linear trend, $p=0.344$).

As discussed in the previous chapter, correlations of operative improvements between different parameters are often poor. A correlation of pre-operative pain analogue scores with walking times and with pain subsections on other scales (table 6.3A) together with correlations of the operative changes in all these parameters, reveal a similar picture (table 6.3B). Neck pain and arm pain correlated well with each other and the operative changes also correlated strongly. While there were good correlations between the analogue scores and SF36 bodily pain score pre-operatively, there were no major correlations of operative change with any of the other scales. Correlations of pain analogue scores and walking times were poor throughout.

There was thus no evidence to indicate that the walking parameter benefits are at the expense of worsened neck pain or arm pain. In contrast, they also appear to improve from surgery, although in a manner uncorrelated with the degree of walking improvement. This lack of correlation means that different patients are improving to different extents in walking versus pain, and thus suggests that it might be useful to record pain in addition to walking in order to gain a fuller picture of overall operative benefit.

A. Pre-op	**Neck Pain**	**Arm Pain**	**EMS Pain**	**SF36 Pain**	**Walk**
Neck Pain	1	-	-	-	-
Arm Pain	0.75(p<0.001)	1	-	-	-
EMS Pain	0.25 (p=0.08)	0.34 (p=0.016)	1	-	-
SF36 Pain	0.50 (p<0.001)	0.58 (p<0.001)	0.37 (p=0.007)	1	-
Walk	0.22 (p=0.125)	0.26 (p=0.065)	0.22(p=0.0.13)	0.20 (p=0.16)	1

B. **Change 6m to 3 y**	**Neck Pain Change**	**Arm Pain Change**	**EMS Pain Change**	**SF36 Pain Change**	**Walk Change**
Neck Pain Change	1	-	-	-	-
Arm Pain Change	0.62(p<0.001)	1	-	-	-
EMS Pain Change	0.04 (p=0.78)	0.19 (p=0.20)	1	-	-
SF36 Pain Change	0.11 (p=0.46)	0.32 (p=0.024)	0.37 (p=0.008)	1	-
Walk Change	0.082 (p=0.57)	0.049 (p=0.74)	0.14(p=0.34)	0.03 (p=0.86)	1

Table.6.3 Correlation of pre-operative pain scores with walking and pain subsection of other scales (6.3A) and the correlation for operative changes (6.3B)

6.5 Discussion

The main finding of the study described in this chapter is that surgery provides a significant and ongoing benefit for CSM patients. This benefit is likely to be not only statistically significant but clinically meaningful if one considers the magnitude of the mean benefit and the fact that this benefit was in a 30-m walking test incorporating a turn, a function likely to be relevant for everyday life.

In contrast, a significant deterioration was found in a group of unoperated patients with CSM over the three years, indicating that decompression has a major and long-lasting impact upon the natural history of CSM. Moreover, matching unoperated patients with decompressed patients of similar severity directly showed both an early improvement from surgery and a subsequent widening of this improvement over time. Clearly, the surgery has altered these patients' disease in a fundamental way.

This study is likely to underestimate the benefits of surgery because a third of the initially unoperated patient group "escaped" the natural history of their condition by having surgery later on over the subsequent 3 years. This of course cannot be avoided for ethical reasons. Matching of the remaining patients with operated patients of similar initial severity seemed the best and most conservative way to analyse the data.

A minority of unoperated patients (3 out of 22) improved over the three years. This was attributed to conservative treatments such as regular massage and physiotherapy. There was no reason to consider that the matched decompressed patients had better access to such treatments. The 3 patients perhaps constitute the small minority of cases from which anecdotal reports have historically and erroneously been drawn that throw doubt upon the value of decompression in CSM.

Of the patients who were operated upon, no firm conclusions can be drawn about the merits of the different procedures performed. Any interpretation of the non-statistically significant trends that existed must be made with caution, especially since the walking data and the JOA revealed different trends! This highlights the fact that different outcome measures should not be interpreted as equivalent. Possibly, one might infer that different operations have different effects on walking and bladder function or pain, but it is just as likely that trends of difference relate to the small numbers in each procedure category. In any event, complex factors determine the surgeon's choice of procedure - some patients may simply be technically more suitable for certain procedures and judgement as to which is the "best" procedure should perhaps at this time be based upon individual cases rather than *en bloc* studies.

Apart from differences with respect to results of individual procedures, the various assessment scales in general correlated with the findings of walking data. All of the scales recorded an early and sustained improvement from surgery. Two scales, the Nurick and the Ranawat, recorded a late deterioration at 3 years (although still better than preoperatively), but this may have reflected the limitations of the scales. For example, with increasing likelihood of assessment by physical and occupational therapists over time, patients may start to use a walking stick, making them worse on the grading systems but perhaps improving their mobility and walking time.

The relative advantages and disadvantages of using a measure that focuses on a single aspect of disability were discussed in the previous chapter. It was concluded that the walking test does not necessarily suffer through measuring only this single aspect. Nevertheless, there may be a minority of patients who mainly have non-walking features

that behave differently. For example, radicular problems or local neck pain might be worsened by surgery. However, the pain analogue score performed in this study revealed that such factors in fact also improve with surgery. Regarding bladder dysfunction, too few patients in this largely *chronic* onset CSM study complained of this symptom to draw any firm conclusions. In any event, bladder dysfunction is rather difficult to quantify and ascribe with certainty to myelopathic pathology.

The quantitative nature of the walking data allowed a correction to be made for age. There is no "normal" value for walking, but only a normal range. It has been found that correcting for the *normal* deterioration in walking with age is possible and when this correction was applied, the findings were not materially altered. However, in addition to the *normal* effect of age on walking there might also be an effect of age on the disease process. Older patients may have a different natural history, tend to do better or worse from intervention, or may simply suffer greater disability from a certain degree of spinal cord damage due to a lack of "reserve". In any event, initial reports by Nurick (Nurick, 1972) on natural history indicated older patients had a more aggressive natural history – an indication to operate. However, other studies have indicated that the older subpopulation (60 +) does not benefit from decompression (Naderi, Ozgen et al., 1998). To investigate this more closely, the present study (where many of the patients were 60 +) uses linear regression analysis to look at the influence of age both on natural history *and* on early and late changes from surgery. Older patients are clearly found to have a worsened natural history if left without intervention. A patient 10 years older would appear to suffer on average an *extra* 6.7 s deterioration of walking time after 3 years. On the other hand older patients did just as well from surgery, when assessed at 6 months as well as when assessed at 3 years after their operation. The study therefore takes the original view of Nurick (Nurick, 1972) and proposes that there is no justification for an "ageist" policy on decompressive surgery.

Finally, the other statement put forward by (Nurick, 1972), that decompression is mainly beneficial for milder disease, was investigated. Here, this study must take the opposite view. Regression analysis of improvement on initial severity indicates that

originally worse patients do better in terms of absolute improvement in walking time and even in terms of relative improvement. The study does not specifically answer the question regarding timing of surgery because it appears that patients at all stages of disease severity will benefit. However, there is certainly no evidence for harm in delaying surgery somewhat. This finding might hopefully help to resolve one of the clinical difficulties faced in management of CSM patients, namely the impression that over prolonged follow up, patients seen to blend imperceptibly from the category of "too mild to accept the risks of surgery" to that of "too late to make a difference". The present data suggest a change in attitude to one where more severely affected patients may be considered to have more room for improvement following surgical intervention. However, one must also consider the case mix involved; the majority of the patients in this study were able to walk, albeit in some cases very slowly. Another population where all the patients at the severe end of the spectrum were already wheelchair bound might reveal different results. It might be better therefore to consider that "if patients are still able to walk, then they have a good prospect of post-operative improvement".

7. Clinical and Radiological Correlates of Severity and Surgical Outcome in Cervical Spondylosis

7.1 Summary

The aim of this study was to determine if radiological features predicted outcome in cervical spondylotic myelopathy (CSM). In a group of 69 patients referred to the National Hospital, Queen Square, London for decompressive surgery, radiographic data from cervical spine MRI scans taken shortly before surgery were analysed by two blinded radiologists each on two separate occasions. The parameters determined were signal change, and presence and severity of compression. Clinical outcome was determined by pre- and post-operative timed walks, MDI, Ranawat and Nurick scores.

There was good inter- and intra-observer reliability for determination of radiological data. A significant relationship existed between MRI signal change and surgical outcome, as measured by improvement in walking parameters. However, this was confounded by the fact that signal change also related to pre-operative walking parameters and that those patients with worse preoperative walking had greater improvement in walking. The relationship between walking data and severity or extent of cord compression was less clear.

It is concluded that cervical cord compression and intrinsic signal change on MRI correlate with clinical severity and, in this population, the presence of signal change is correlated with greater operative improvement following surgery. However, confounding factors and the lack of strength of correlation mean that these radiological measurements are insufficient to be used as a reliable tool for predicting benefit from surgery in individual patients.

7.2 Introduction

The previous chapters of this thesis, as well as previous studies(Chiles, Leonard et al., 1999;Hirabayashi, Toyama et al., 1999;Naderi, Ozgen, Pamir, Ozek, and Erzen, 1998), have indicated that surgical decompression is largely an effective means of

treatment for cervical spondylotic myelopathy (CSM). However, there are still individual patients that tend to do badly from surgery. While many clinical parameters have been used to attempt to predict good benefit from surgical decompression (Emery, Bohlman et al., 1998), many are unreliable and all in general are applicable only to populations rather than also to individuals.

Previous studies have been conducted on a retrospective basis comparing radiographic findings with clinical severity and with surgical outcome. Perhaps it is possible to use radiographic data as an adjunct to clinical data not only to make the diagnosis of CSM but also to predict which individuals are likely to receive best results from surgery and perhaps which ones should not be operated upon at all.

As expected, it has been clear from early post-mortem studies that the severity of myelopathy relates to the severity of cord compression as measured by antero-posterior cord diameter (Ogino, Tada, Okada, Yonenobu, Yamamoto, Ono, and Namiki, 1983). This has been borne out by similar MRI measurements of the cervical cord in life (Wada, Yonenobu et al., 1999). Some retrospective studies have also shown that, as well as correlating with clinical severity, more severe cord compression as measured by cord diameter (Bucciero, Vizioli et al., 1993) or by transverse area (Takahashi, Sakamoto et al., 1987) is associated with a poorer post-operative outcome. However, such findings may depend on the criteria used to measure outcome. For example, the latter study solely employed a retrospectively determined Japanese Orthopaedic Association (JOA) score (Hirabayashi, Toyama, and Chiba, 1999) converted into the scale of relative recovery.

The aetiology of CSM does not necessarily relate only to direct compression, i.e. pinching between osteophytes and disc material anteriorly and the posterior longitudinal ligament posteriorly. Subclinical compression may result in chronic cord changes arising from alteration of the cord microcirculation (al-Mefty, Harkey et al., 1993), while mechanical stretching of the cord over spondylotic ridges during neck flexion has also been cited as important(Breig, Turnbull, and Hassler, 1966). Thus, in addition to measurement of antero-posterior diamaters at the compression site, other radiological parameters, such as intrinsic signal change, may also be useful as indicators of CSM.

Takahashi and colleagues (Takahashi, Sakamoto, Miyawaki, and Bussaka, 1987), found increased intrinsic signal on T2 and proton density imaging in a proportion of patients with compressive lesions of the cervical canal, and such changes were more likely when there was disc herniation or ossification of the posterior longitudinal ligament. Subsequent studies on the use of signal change to predict outcome are conflicting. Okada et al (Okada, Ikata et al., 1993), found that signal change correlated with a poorer postoperative outcome while Morio et al found no correlation. Wada et al (Wada, Yonenobu, Suzuki, Kanazawa, and Ochi, 1999), found a correlation only in as much as multi-level signal change was associated with postoperative upper limb radicular problems. Finally, in another retrospective study, Kumar et al (Kumar, Rea et al., 1999), found that *post-operative* signal change had only a weak correlation with postoperative clinical status.

Despite an uneasy consensus that a greater severity of compression and the presence of signal change predict a bad outcome from surgery, common clinical practice when compression is mild is actually to delay operative decompression until the development of significant symptoms. Instead, early surgery is often offered to patients with more severe compression or with clear signal change.

Many of the difficulties in interpretation of studies on clinico-radiological correlates and in relating such studies to clinical practice may lie in the perception that investigators' findings are very dependent upon the way clinical outcome has been measured, and that such clinical measures are not perceived to be applicable to individual patients. The properties of timed walk parameters that have been explored in the present series of studies make them particularly useful for assessment of operative change in CSM and also enable more powerful statistical analysis (Dun-Rankin, 1983;Casey, Bland, and Crockard, 1996) and more powerful correlative analysis with other data.

The present study therefore employs prospective pre- and post-operative measurement of walking data to attempt to resolve conflicts in the literature and in clinical practice over the way radiological cord compression and signal change may predict operative outcome. Specifically, it was first determined if basic radiological

parameters of compression and signal change can be measured in a reliable and reproducible way by different radiologists. Second, the MRI findings that correlated best with clinical severity were explored. Finally, the possibility that pre-operative MRI findings could reliably predict the outcome of surgery was examined.

7.3 Methods

Radiographic correlations were made on 69 patients, some of whose data have already been described in previous chapters.

7.3.1 Clinical Functional Assessment

This was performed as usual by pre- and post-operative 30-m timed walks and application of CSM severity scales; values for the (MDI), the Nurick functional walking score and the Ranawat impairment score were collected by direct questioning.

7.3.2 Radiographic Assessment

Radiographic data from cervical MRI scans taken shortly before surgery were analysed independently by two experienced Neuroradiologists each on two separate occasions about 3 months apart (fig. 8.1 MRI of mild, moderate, severe compression). They were blinded to the clinical details and the scans were presented in random order. The parameters determined were:

i) The presence of T2-weighted signal change at each cervical segmental level handled simply as a yes/no response.

ii) The presence and severity of cord compression at each level, with compression severity was rated as none=0, mild (flattening or concavity of the anterior surface only)=1, moderate (up to 50% reduction in maximal sagittal diameter)=2 and severe (over 50% reduction in sagittal diameter)=3, (fig. 7.1). Total compression was a parameter derived by summing the compression scores at all spinal levels in an individual patient.

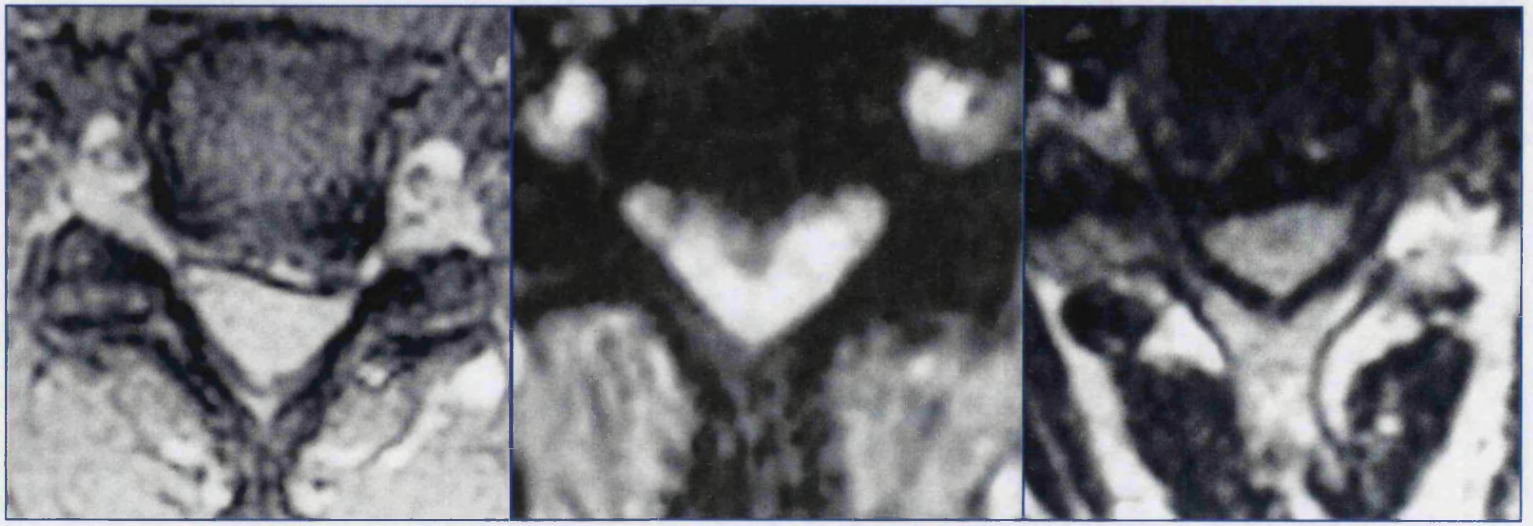


Fig. 7.1 MRI scan showing severities of mild (a), moderate (b) and severe (c) as rated by the two radiologists.

7.3.3 Data Analysis

Exploratory correlation coefficients were first performed between all the parameters. Since the walking data alone were continuous and approximately normally distributed, these data alone could be analysed in a much more powerful way by analysis of variance (ANOVA), segregating groups according to different scores on the discontinuous radiographic data. Where there were more than two possible scores for a radiographic factor (e.g. zero, mild, moderate, severe severity), trends and contrasts were determined to see at which level of a factor, if any, there was a cut-off between different groups (e.g. mild to moderate versus severe or mild versus moderate to severe).

7.4 Results

In all, 69 patients were studied, consisting of 46 males (mean age 57.8 ± 13.1 (SD)) and 23 females (mean age 62.52 ± 13.8 (SD)). There was one death 3 months following surgery due to cardio-respiratory failure and one patient had a corneal abrasion from her theatre eye pads. Two patients had wound infection. The following types of surgical procedures were carried out: 28 Smith Robinsons, 7 Clowards, 13 laminectomies, of which 2 had lateral mass plates inserted, and 21 laminoplasties with mini-plating.

7.4.1 Walking Data

There were 4 patients who were unable to walk at all preoperatively and were able to do so post-operatively. There were two who could walk preoperatively and who became unable to do so following surgery. In addition, there was one other patient who was unable to walk either before or after surgery. Of the remainder (62), the mean preoperative walking time was 62.7 s ± 7.2 (SE) and that postoperatively was 45.7 s ± 4.6 (SE). This represented a highly significant improvement ($p<0.001$; 2 tailed paired sample t-test assuming unequal variances). The number of steps taken during the walk was 68.3 ± 3.8 (SE) preoperatively and 57.2 ± 2.7 (SE) postoperatively. This was also a significant improvement ($p<0.001$; 2 tailed paired sample t-test assuming unequal variances) (fig. 7.2).

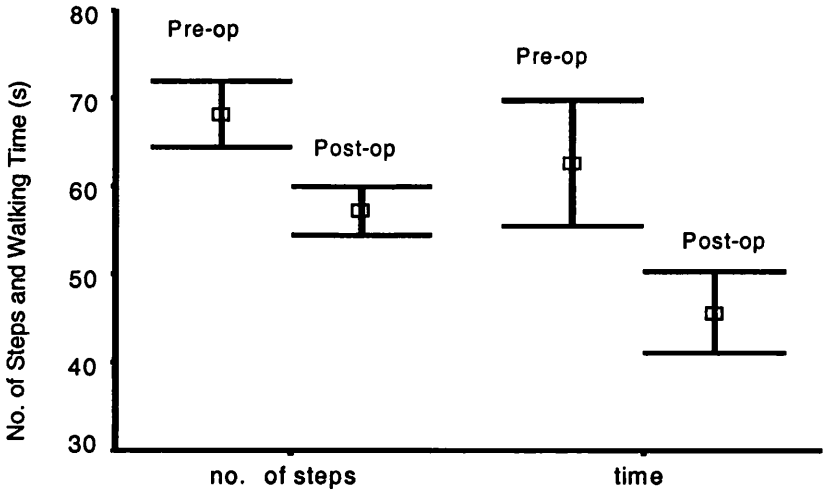


Figure 7.2. Walking times for 30-m timed walks. There was a highly significant improvement (2 tailed paired sample t-test assuming unequal variances) in the number of steps taken ($P<0.001$) and walking time ($p<0.001$) following surgery.

7.4.2 Radiographic Data

7.4.2.1 Reliability: There was an extremely high level of reliability of signal change estimation between analysis sessions for each radiologist and a rather lower reliability between radiologists both for the first and second sessions. For example, for signal change, there was one disagreement within radiologists (99.3% correct) and 7 between radiologists (94.9% correct). The full reliability results are shown in table 7.1. There was also reasonable, though not as good, reliability for estimation of the number of levels of cord compression and compression severity (mild, moderate, severe).

	Intra-rater 1	**Intra-rater 2**	**Inter-rater 1**	**Inter-rater 2**	**Overall**
Levels of signal change	0.99	1.0	0.94	0.95	0.98
Levels of compression	0.98	0.98	0.82	0.84	0.93
Worst compression severity	0.96	0.97	0.80	0.89	0.94

Table 7.1. Reliability scores comparing the results of the two radiologists each on two sessions. A score of 1 indicates a perfect match between sessions 1 and 2 or between raters 1 and 2. All reliabilities were reasonable, with intra-rater reliabilities being better than between raters and signal change being most reliably assessed.

For subsequent analysis, where there was a single disagreement in value over the four estimates of each parameter within and between radiologists, the mode value was taken. If there were two or more (out of a possible four) disagreements, cases were excluded from analysis; in this way 2 cases were excluded from analysis of signal change, 13 from analysis of the number of compression levels and 12 cases excluded from analysis of compression severity. Since disagreements occurred in interpretation of mild as well as severe compression, such exclusions were not considered to introduce bias but they would reduce statistical power.

7.4.2.2 Range of values

For each of the above radiographic parameters, there was a reasonable range of number of levels and of severity. In other words, the data did not tend to cluster at one particular value. This property is vital for the use of radiographic data in determination of severity - there must clearly be a wide range to distinguish different patients properly. A parameter whose scores were identical for every patient would be virtually useless.

7.4.3 Correlation Coefficients

As an initial exploration of correlations between the various parameters, Spearman rank correlations were performed (table 7.2 compares walking and radiographic data and table 7.3 compares the functional scales (MDI, Nurick and Ranawat) with the same radiographic data).

Some interesting positive correlations were apparent. Some, like the effect of age, have been discussed in previous chapters. Regarding radiographic correlations, older patients tended to have more levels of cord compression but the severity at each of these levels was not more severe. In contrast to age, no sex correlations were apparent.

Within radiographic data, there were modest but significant correlations between the presence and number of levels of signal change and the severity of compression, but less so with the number of levels of compression(Singh, 2001).

Correlations between the walking and the radiographic data were modest at best. There was no particular increase in correlation by using the derived parameter of total compression (the summed severities of compression at different levels).

(One patient had no clear signal change or cord compression. Instead this patient had been operated upon for minor multilevel spondylotic changes. His initial walking time was 18 s with 44 steps taken and following surgery there was almost no change. Thus, he followed the general positive correlation found in this study that mildly radiographically affected patients have less benefit – if one is initially normal, there is no benefit to be had!)

	Age	Sex	Walking time	Steps taken	Change in walking time	Change in steps taken	Levels of signal change	Levels of compress-ion	Compression severity	Total compress-ion
Age	-	0.17 *P*=0.16	**0.52** ***P*<0.001**	**0.52** ***P*<0.001**	0.23 *P*=0.047	0.047 *P*=0.69	0.10 *P*=0.41	0.30 *P*=0.012	-0.046 *P*=0.71	0.27 *P*=0.026
Sex		-	-0.006 *P*=0.96	0.10 p=0.375	-0.037 *P*=0.76	0.005 *P*=0.97	-0.15 *P*=0.21	0.05 *P*=0.67	-0.065 *P*=0.60	0.0078 *P*=0.95
Walking time			-	**0.79** ***P*<0.001**	**0.52** ***P*<0.001**	**0.36** ***P*=0.002**	0.18 *P*=0.14	0.27 *P*=0.024	-0.022 *P*=0.85	0.23 *P*=0.063
Steps taken				-	**0.54** ***P*<0.001**	**0.49** ***P*<0.001**	0.05 *P*=0.69	0.24 *P*=0.045	0.13 *P*=0.29	0.27 *P*=0.025
Change in walking time					-	**0.79** ***P*<0.001**	0.17 *P*=0.16	0.027 *P*=0.82	0.20 *P*=0.10	0.15 *P*=0.23
Change in steps taken						-	0.17 *P*=0.18	-0.012 *P*=0.92	0.20 *P*=0.10	0.089 *P*=0.47
Levels of signal change							-	0.27 *P*=0.025	**0.38** ***P*=0.001**	0.30 *P*=0.011
Levels of compression								-	0.26 *P*=0.03	-
Compressio n severity									-	-
Total compression										-

Table 7.2. Two-tailed Spearman's correlations between walking and radiological parameters. Those values in bold indicate significant correlations allowing for Bonferroni correction.

	MDI	Nurick	Ranawat	MDI change	Nurick change	Ranawat change	Levels of signal change	Levels of compress-ion	Compress-ion severity	Total compress-ion
MDI	-	**0.71 *P*<0.001**	**0.47 *P*<0.001**	**0.65 *P*<0.001**	0.04 *P*=0.76	-0.09 *P*=0.44	0.20 *P*=0.11	0.33 *P*=0.006	0.15 *P*=0.21	**0.38 *P*=0.001**
Nurick		-	**0.66 *P*<0.001**	**0.36 *P*=0.002**	0.22 *P*=0.061	-0.07 *P*=0.55	**0.34 *P*=0.005**	**0.34 *P*=0.004**	0.07 *P*=0.60	0.33 *P*=0.006
Ranawat			-	0.25 *P*=0.038	**0.34 *P*=0.004**	0.29 *P*=0.013	**0.39 *P*=0.001**	**0.35 *P*=0.003**	0.31 *P*=0.01	**0.43 *P*<0.001**
MDI change				-	**0.37 *P*=0.002**	**0.34 *P*=0.003**	**0.33 *P*=0.005**	0.11 *P*=0.36	0.28 *P*=0.02	0.19 *P*=0.13
Nurick change					-	**0.65 *P*<0.001**	0.26 *P*=0.03	-0.06 *P*=0.65	0.006 *P*=0.96	-0.05 *P*=0.71
Ranawat change						-	0.11 *P*=0.36	-0.07 *P*=0.55	0.08 *P*=0.49	-0.11 *P*=0.36

Table 7.3. Two-tailed Spearman's correlations between measurement scales and radiological parameters.

7.4.4 MDI, Nurick and Ranawat Scale Corelations

Correlation coefficients (table 7.3) yielded similar results to those in chapter 6. There was an indication, especially for the MDI scale ($r=0.65$; $p<0.001$), that there was greater post-operative improvement when the initial pre-operative score indicated more severe disease. This was previously found to be the case also for walking data. Functional (MDI and Nurick) and impairment (Ranawat) severity in general correlated with signal change and with the number of levels of cord compression but not with compression severity.

The only significant radiological *predictor for operative benefit*, as determined by correlation coefficients for these scales, was that a greater pre-operative signal change was associated with a greater improvement in MDI following surgery. This is shown by the mild (0.33) but significant ($p=0.005$) correlation between post-operative change in MDI level with the pre-operative number of levels of signal change.

7.4.5 Analysis of Variance and Contrasts

Using non-parametric Spearman correlations for the walking data is likely to underestimate any relationships between radiographic data. The data lend themselves instead to analysis of variance, separating different groups of patients according to their level of severity on various radiographic parameters. Such analysis is much more sensitive to patterns and trends.

7.4.5.1 Signal change (fig 7.3): The presence and number of levels of signal change was found to be significantly greater in those patients who had a worse (longer) initial walking time ($p=0.0011$) (fig. 7.3a) and, most interestingly, in those patients who had a greater *improvement* in walking time following surgery ($p=0.001$) (fig. 7.3b). The same applied for the number of steps taken ($p=0.0048$) (fig. 7.3c) and for the improvement in the number of steps taken ($p=0.0024$) (fig. 7.3d). There was a significant linear trend for walking time ($p=0.013$), for the improvement in walking time following surgery ($p=0.0024$) and in the improvement in number of steps following surgery ($p=0.0044$), but

not quite for the absolute number of steps taken preoperatively (p=0.06). Analysis of contrasts generally revealed no particular contrast between number of levels, although the 3 patients who had three levels affected had particularly severe walking parameters and showed good improvement.

It is possible that the apparent predictive effect of signal change on improvement in walking parameters could reflect that more signal change also relates to worse pre-operative absolute walking time and that these patients simply had more room for improvement. In fact, on linear regression analysis, it was found in this study that worse initial walking times did indeed result in greater *absolute* improvements ($p=1.05 \times 10^{-9}$) but there was only a much weaker trend for greater *relative* improvements (c.f. chapter 4 where the trend for greater relative improvement only just reached significance). When relative improvements instead of absolute improvements in walking data were compared with signal change, there was *no* significant difference for different levels of severity (relative walking time improvement, p=0.361; relative number of steps improvement, p=0.355), suggesting that the assumption may well be true – the relationship between signal change and improvement in walking parameters may perhaps be secondary to the fact that worse affected patients have greater improvements.

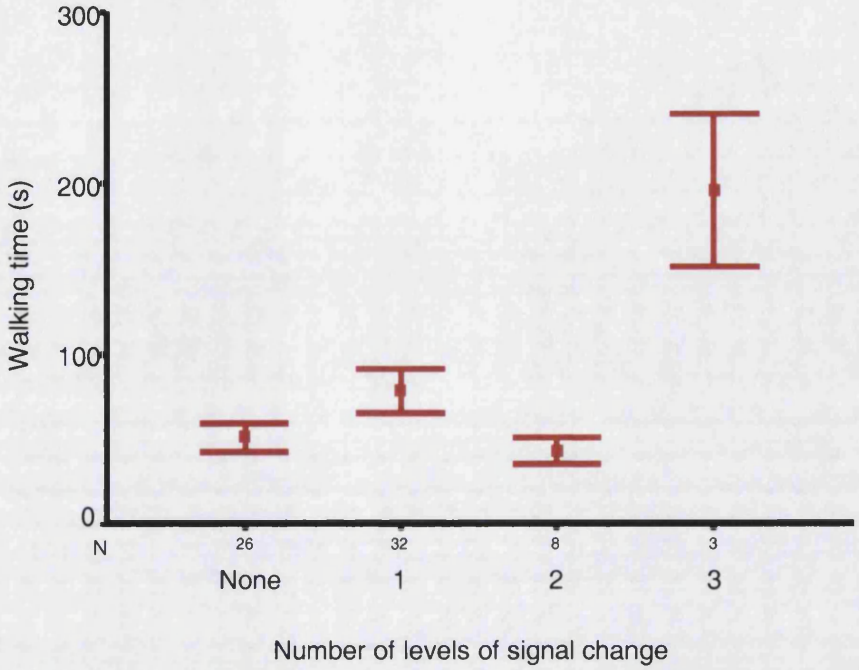


Fig. 7.3a

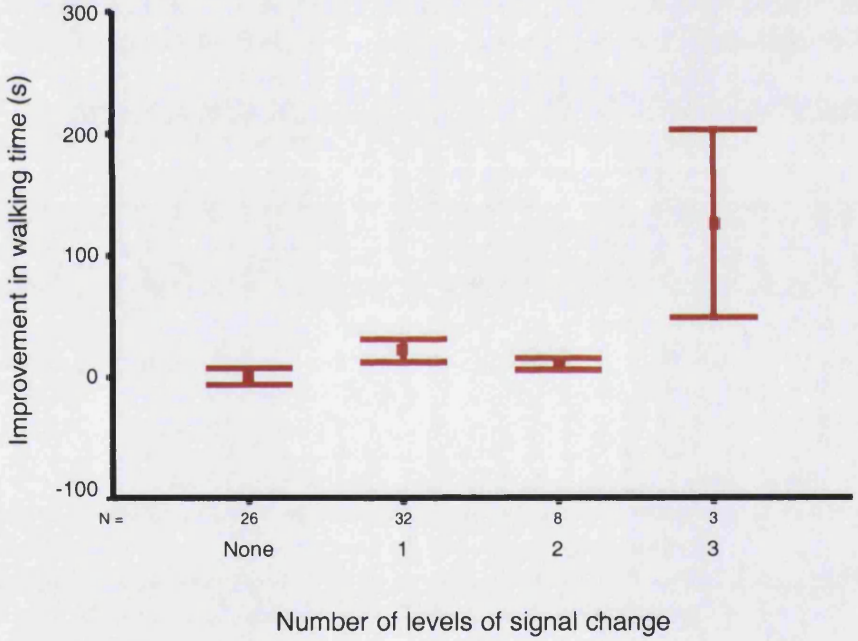


Fig. 7.3b

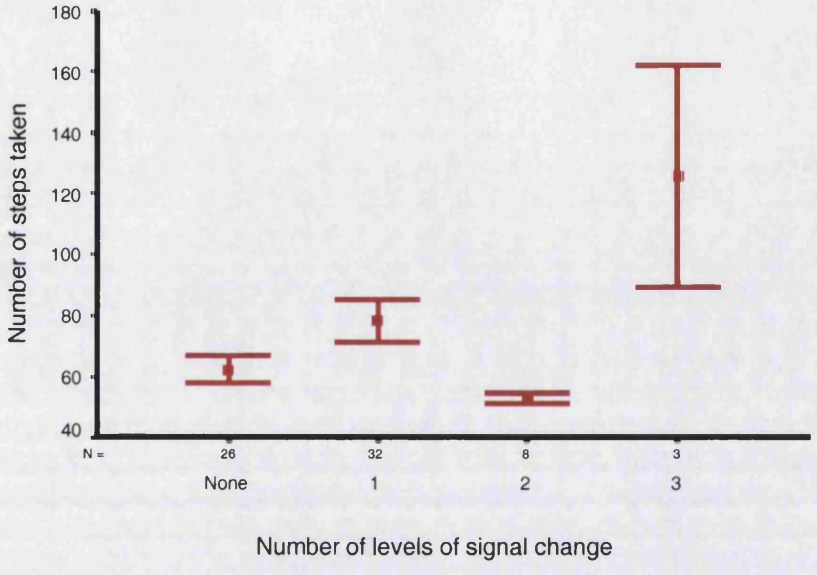


Fig. 7.3c

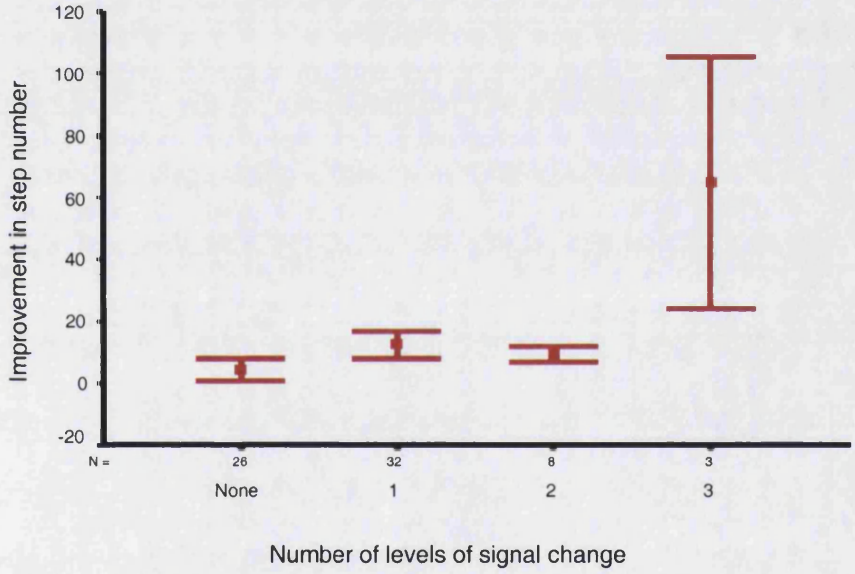


Fig. 7.3d

Fig. 7.3 Pre-operative walking parameters and post-operative change in walking parameters for different numbers of levels of cord MRI signal change. The smaller numbers on the x-axis indicate the number of patients out of 69 in each category. Those patients with more levels of signal change had worse initial walking times (A) and, most interestingly, greater improvement in walking times (B). The same applied for the number of steps taken (C, D). Thus a greater number of levels of signal change predicts greater surgical benefit.

7.4.5.2 Number of levels of compression (fig 7.4): No individual walking time (fig. 7.4a), difference in walking time following surgery (fig. 7.4b), steps taken (fig. 7.4c), or

difference in steps taken (fig. 7.4d) groups were significantly different when segregated according to the number of levels of cord compression. However, there was a significant linear *trend* for walking time (p=0.02) and for steps taken (p=0.03) to be greater with greater number of levels affected. On contrast analysis assuming separate variances, the main significant increase was between one level affected and more than one levels affected (i.e. a combination of the groups of zero and one level versus a combination of the groups with more than one level) (walking time p=0.001; steps taken p=0.006).

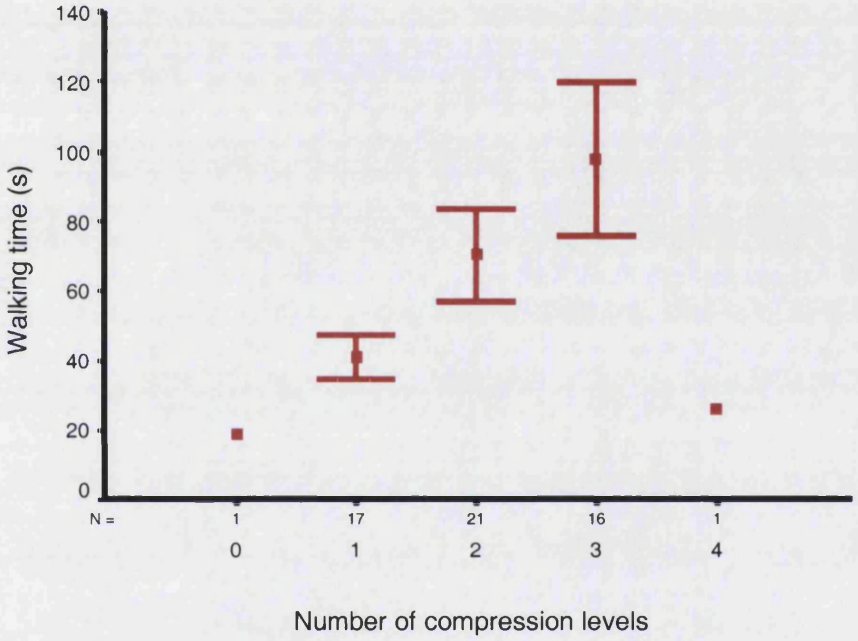


Fig. 7.4a

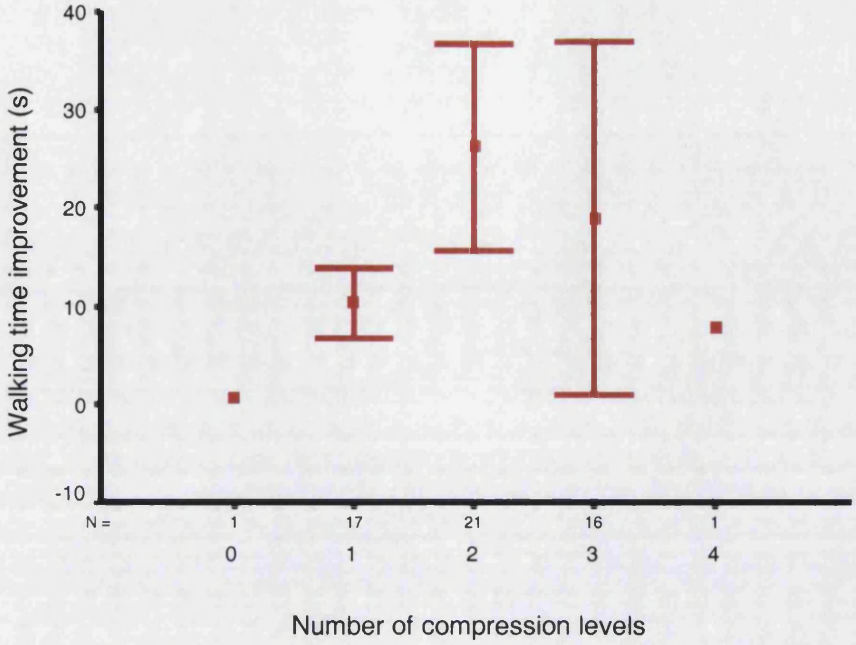


Fig. 7.4b

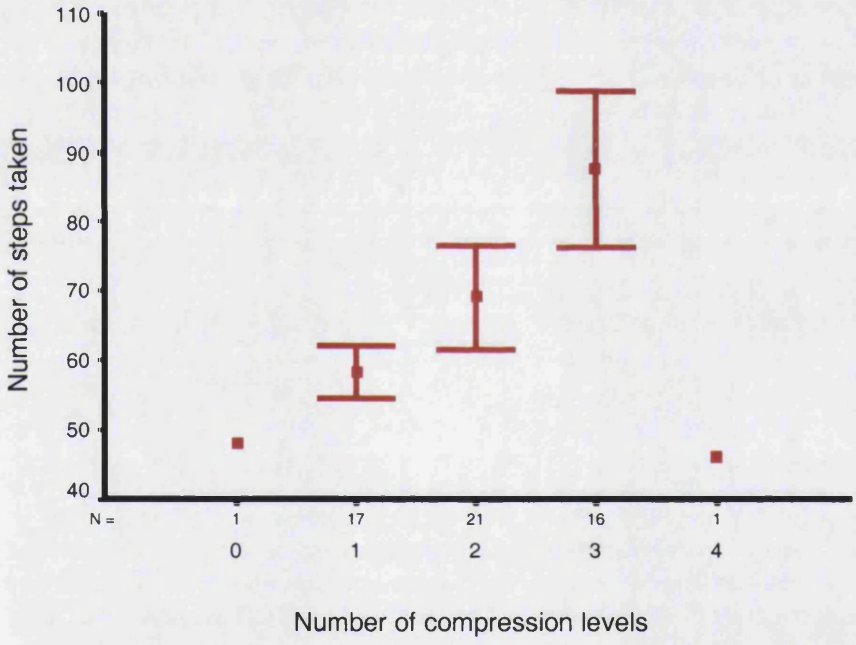


Fig. 7.4c

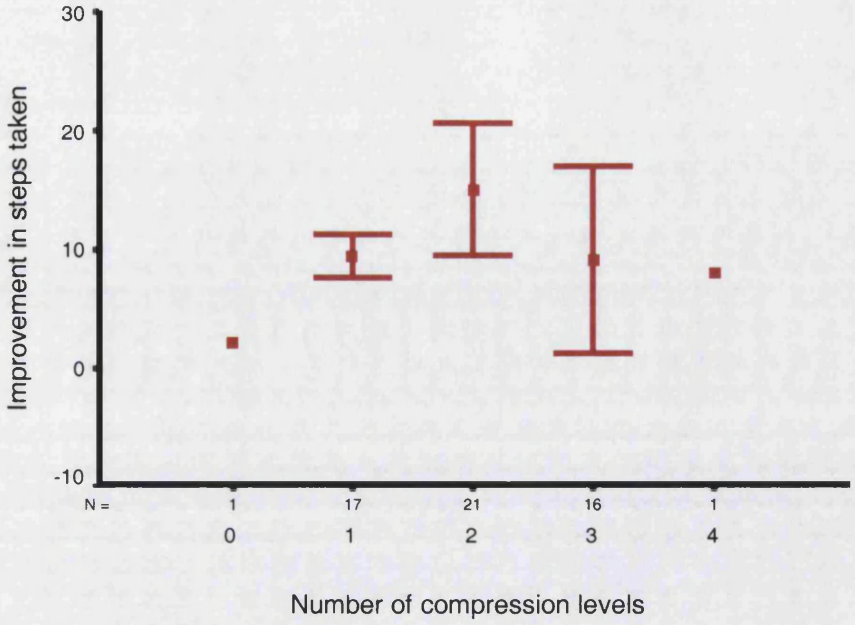


Fig. 7.4d

Fig. 7.4. Walking parameters and post-operative change in walking parameters for different numbers of levels of cervical cord compression on MRI. Walking times were not significantly different between patients with different number of compression levels (A), although there was a significant linear trend for worse walking times when more levels of compression. There was no relationship between improvement in walking time following surgery and the number of compression levels (B). Similar results applied to the steps taken.

7.4.5.3 Compression severity (fig.7.5): On looking at compression severity at the worst site, there were again no significant differences between groups for any of the four walking parameters (figs. 7.5a-7.5d), nor were there any significant linear trends. Analysis of contrasts revealed the mild groups to have better walking times ($p=0.001$) and number of steps ($p<0.001$) than moderate to severe groups, and for postoperative improvements in walking time ($p=0.002$) and steps taken ($p=0.006$) to be poorer for the mild group when contrasted with a combination of the moderate to severe groups.

On looking at a total measure of cord compression in each patient (i.e. summing the severities at the different sites so that, for example, one mild level of compression (severity 1) together with one severe level of compression (severity 3) in the same scan gives a total severity of 4), there were significant linear trends for walking time ($p=0.007$) and steps taken ($p=0.005$).

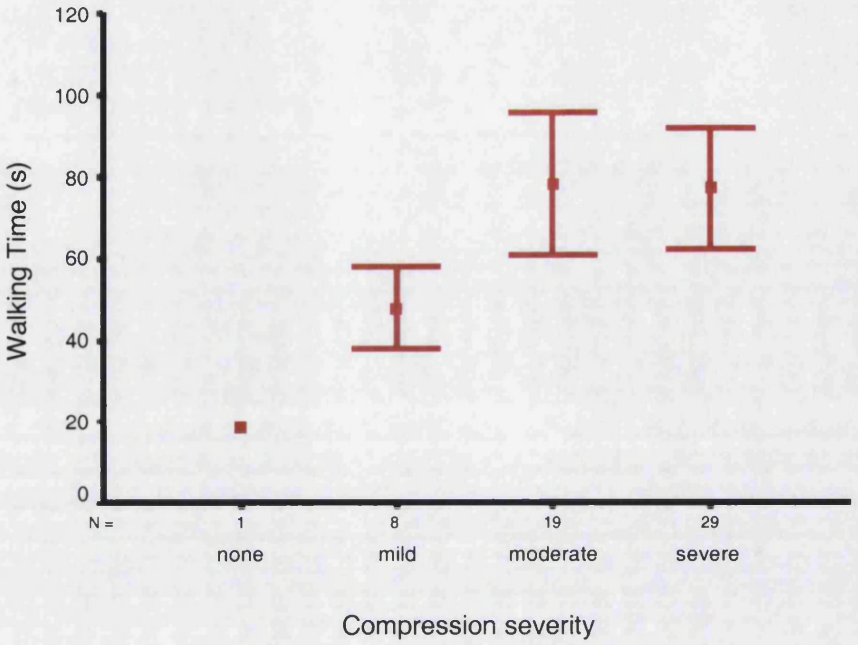


Fig. 7.5a

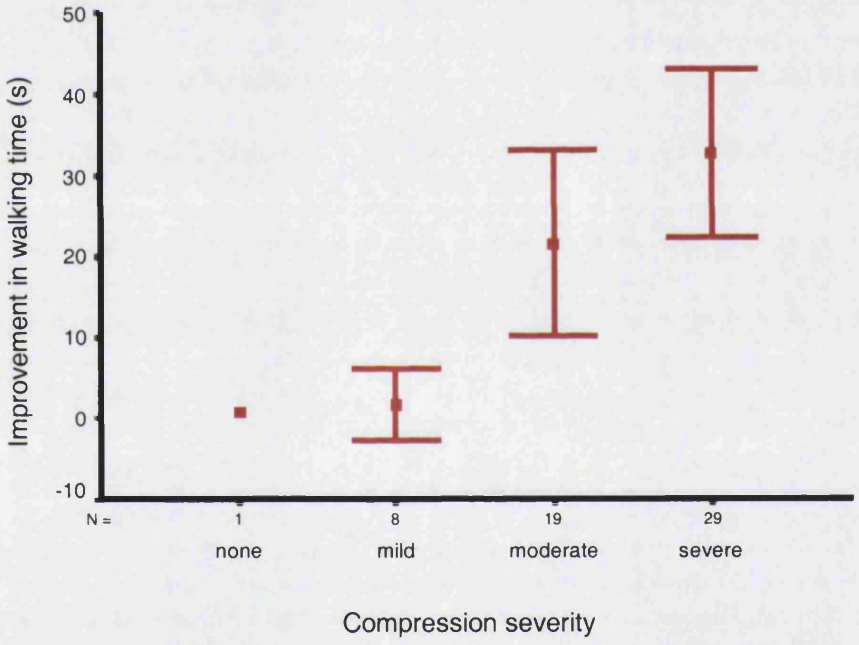


Fig. 7.5b

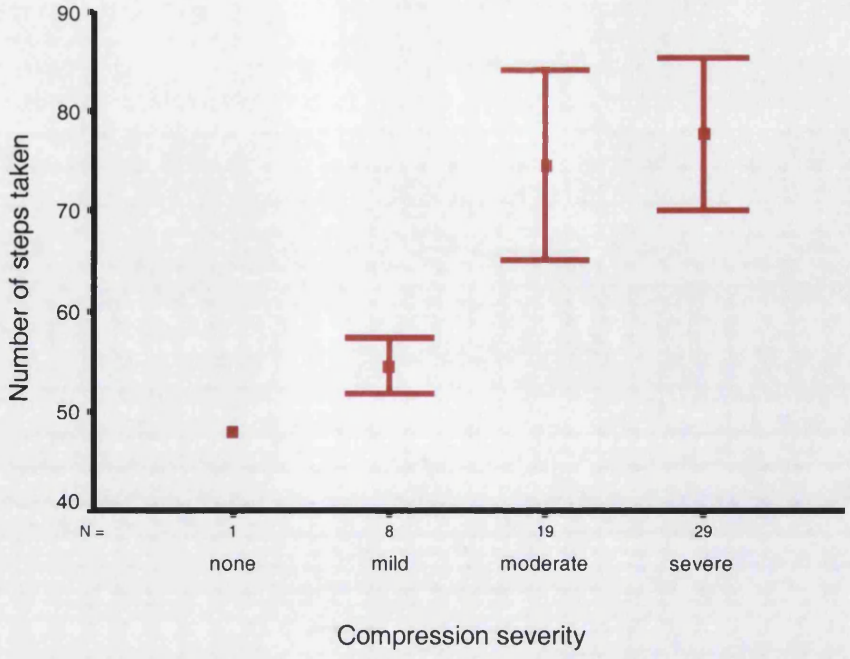


Fig. 7.5c

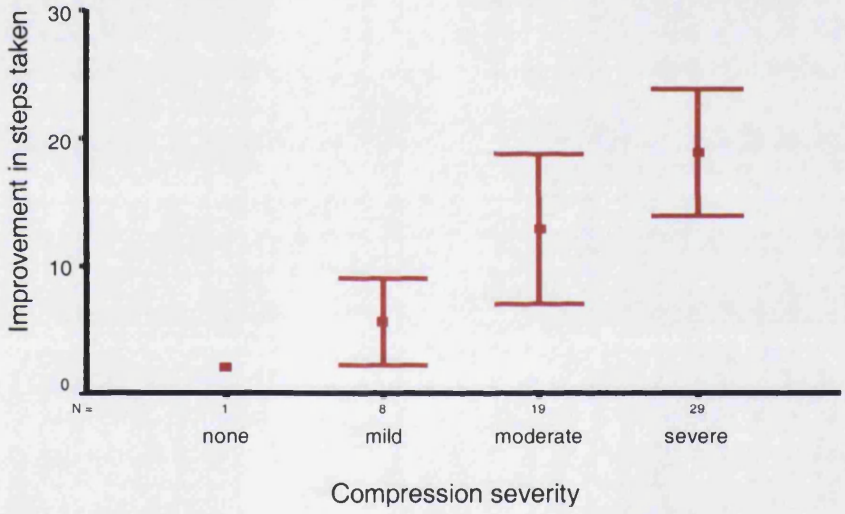


Fig. 7.5d

Fig. 7.5 Walking parameters and post-operative change in walking parameters for different degrees of cord compression (none, mild, moderate or severe compression at the worst level if there were multiple levels). There were no significant differences in walking parameters or post-operative improvements in walking parameters for different degrees of compression, nor were there any significant linear trends. However there were significant contrasts between mild v/s moderate to severe compression for walking time (A), improvement in walking time (B), number of steps (C), and improvement in number of steps (D).

7.5 Discussion

Radiographic Parameters

Although a number of previous studies (Fukushima, Ikata et al., 1991;Hamburger, Buttner et al., 1997;Takahashi, Sakamoto, Miyawaki, and Bussaka, 1987;Wada, Yonenobu, Suzuki, Kanazawa, and Ochi, 1999;Yu, du et al., 1986) have already looked at radiographic parameters in cervical spondylotic myelopathy, this study has formally analysed the reliability, sensitivity and likely validity of such data prospectively in 69 CSM patients.

Reliability analysis reveals a very high consistency in determination of pre-operative T2-weighted MRI signal change and cord compression when performed by the same rater at different times. However, assessment of compression (none, mild, moderate or severe) showed poorer reliability between different raters. Although the actual reliability scores remained high, the limitation of use of compression assessment in this manner is illustrated by the fact that 12 or 13 cases had to be removed from subsequent analysis because of disagreement between raters.

It is desirable that any method of grading be sensitive and this can be gauged by determining the range of grades that are obtained in the population of interest. A good range is present for the number of levels of signal change and compression and for the severity of compression, indicating that the grading system is sensitive in distinguishing patients with "mild" and "severe" scans. Since all the patients in the present study ultimately had surgery, no helpful conclusions could have been drawn if all the patients were uniformly graded as "severe". The clinical parameters in fact indicate that our patients had a large variation in clinical severity and a wide overlap in walking parameters and rating scores with other patients choosing not to undergo surgery. Thus, despite the fact that no patients were actually treated conservatively, our population was still suitable for testing parameters attempting to predict operative benefit.

Scales and grading systems are generally validated by comparing with already established systems. Since the radiographic data purport to relate to myelopathic severity, a suitable comparison may be made with already-validated clinical scales of myelopathic

severity. Significant correlation exists between the pre-operative Nurick and Ranawat scales and the number of levels of signal change and compression but not compression severity. More powerful analysis of variance of walking data indicates that patients with more levels of signal change do have worse walking parameters and a trend for a similar relationship with respect to levels of compression, but not for compression severity. Thus, in general, there is some support for the fact that signal change in particular is a valid measure of myelopathic severity.

Radiographic Parameters as Predictors of Surgical Benefit

Signal change proved to be the clearest predictor of surgical benefit; patients whose scans had more levels of signal change subsequently had greater benefit from surgery as measured by improvement in walking parameters. The findings are strengthened by significant linear trends indicating that the greater the number of levels of signal change, the worse the initial walking performance and the greater the operative benefit. (There was no particular contrast or cut-off point between particular numbers of levels of signal change.) There was less strong statistical evidence that a greater number of compression levels or greater severity of compression predicted greater post-operative improvement.

The finding that *worse* signal change predicts *greater* surgical improvement is clearly at variance with much of the existing literature. Some previous studies (eg(Wada, Yonenobu, Suzuki, Kanazawa, and Ochi, 1999),) have converted absolute changes on *non-ratio* functional rating scales into changes relative to starting level before analysis. The present study reveals that conversion to relative improvement removes the association between signal change and improvement; this relates to the fact that improvement is itself greater in patients with more severely affected walking pre-operatively. Thus the particular way data are handled may be crucial in determining the conclusion drawn.

In addition to the way rating scale data are analysed, differences between the present and previous studies might initially be thought also to reflect a fundamental difference between the scales themselves and walking data as measures of clinical

assessment. However, in this study three ratings scales (MDI, Nurick and Ranawat) were used in addition to the walking data and these also had a positive correlation between worse signal change and greater improvement. The correlation between MDI improvement and signal change (r=0.33) reached significance (p=0.005), but the other correlations were weak. Perhaps, in a small retrospective study using only one rating scale, which as we have discussed allows less powerful statistical analysis than is possible for continuously variable walking data, there is a danger of an erroneous result. When one considers that signal change may reflect oedema, and thus reversible damage, as well as gliosis or myelomalacia (Kaiser and Holland, 1998), the present study's finding that such change correlates with greater clinical improvement is not unexpected.

Conclusions

Despite the predictive correlation of worse signal change for greater benefit from surgery, there remains considerable overlap between radiological severity groups so that the finding largely applies to populations rather than directly to practice in individual patients. For instance, it can be seen (fig. 7.4) that a significant number of patients will improve following surgery despite having no MRI signal change. However, important conclusions for practice in individual patients may still be drawn from this study. The presence of signal change should certainly not contraindicate operative decompression, but if anything lead to optimism over outcome. Nevertheless, while MRI scanning of the cervical spine is an invaluable tool for diagnosis and operative planning, it is clear that the basic radiographic features of cord compression or signal change cannot alone predict surgical outcome in individual patients(Singh, 2001). Clinical as well as radiographic criteria should be used to guide decisions regarding whether or not to operate in CSM.

8. General Discussion and Conclusion

As has been described in detail, existing quantitative measures of assessment of severity of cervical spondylotic myelopathy appear not to meet the needs of the consensus of surgeons and other health workers involved in the care of such patients. Different studies in the literature have employed different measures and, as a result, attempts to gain an overall perspective on the efficacy of surgical treatment for CSM have been very conservative. Rowland (Rowland, 1992) was only able to quote a 50:50 chance of benefit from surgery and a 15% + chance for deterioration.

After first gaining a perspective on clinicians' attitudes to and use of assessment scales in CSM, the initial aim of the studies constituting this thesis has been for the first time to make direct comparisons of all the most commonly used scales on the same patients. In particular, it was considered important to look at the way the scales recorded operative changes rather than just severity at a single point in time. This is because the scales will primarily be used to assess changes over time and the effects of intervention, not simply to compare different patients' disease severity. It was found that all the scales examined (JOA, EMS, MDI, Nurick, Ranawat, SF36, Odom's criteria) were able to record clear operative changes, but in other respects they have properties that made them more or less useful. Such properties include absolute sensitivity, sensitivity to change, reliability/ internal consistency and validity.

For example, in the group of patients operated upon in this study, despite recording operative change, the poor absolute sensitivity of the Ranawat scale means that, if used alone, it would have proved useless for scales' secondary function of distinguishing levels of severity. Thus it would have been impossible to look meaningfully at the effects of initial severity on operative benefit, or to correlate severity with other criteria such as radiographic features.

In this comparison of scales, it could be demonstrated that the MDI scale, adapted from the Activities of Daily Living (ADL) and previously validated only for rheumatoid myelopathy (Casey, Bland, and Crockard, 1996), is valid to use for cervical spondylotic

disease in general, and its properties compare rather favourably with other scales such as the JOA measure (Singh and Crockard, 2001).

The SF36 measure showed initial operative improvements in all categories and so is also likely to be a valid measure in CSM. It has the advantage of being more rigorously designed to yield interval rather than merely ordinal scale behaviour, so that the expression of percentage of normal is a meaningful one. However, in the experience of this study, patients found it tedious to complete and would have been happier with a more "concrete" measure of their performance. Nevertheless, an important aspect of the SF36 in the modern climate of financial justification of medical procedures is that it provides an important Government-recognised measure of relative handicap change from intervention. This can be converted into a quality of life measure (e.g. QALYs) to estimate cost effectiveness. The conclusion that would be drawn from these studies is that the SF36 should be used as a vital secondary outcome measure in research trials of intervention but is unlikely to find itself used in daily clinical practice.

The second goal of the thesis was to explore the use of the walking test in assessment of CSM. Walking tests over 10-m have been validated in previous studies by correlation with Functional Ambulation Categories (Holden, Gill, Magliozzi, Nathan, and Piehl, 1984), extent of personal mobility(May, Nayak et al., 1985) as well as with clinical assessment of gait pattern (Wade, Wood, Heller, Maggs, and Langton, 1987) and leg strength (Goldfarb and Simon, 1984). The detailed gait pattern analysis of a small number of patients analysed on a treadmill has already been shown to reveal quantifiable abnormalities in myelopathy and to record improvement in these abnormalities following decompressive surgery (Kuhtz, Johnk, Mader, Stolze, and Mehdorn, 1999). However, in this thesis, practicality and ease of use were considered a priority and it was felt that a simple timed walk might be the most convenient measure of walking. A 30-m walk with a turn in the middle was chosen to provide a realistic model of the kind of walking that reflects disability. The length tests endurance to a certain extent and the turn tests manoeuvrability. However, the test should be easy enough that most CSM patients are able to complete it, albeit slowly.

It was indeed found that the walking test sensitively recorded severity level and changes resulting from surgery. In a second larger study comparing the test directly with previous measures, its properties were rather favourable. Of the two aspects of the walking test, walking time had advantages over the number of steps taken.

The final goal of the thesis was to use the walking test, not simply to explore the properties of different forms of quantification of CSM, but actually to answer important questions about the natural history of CSM and the effects of intervention. For this purpose, a group of patients was followed up over a considerable period (3 years) and compared with a similar group of unoperated myelopathic patients. Important conclusions could be drawn from this study, that do not necessarily correspond with the findings of previous studies. First, patients of a severity range similar to that of the study (an unselected population referred to a major Neuroscience centre, but possibly milder than in some centres where many patients may already be wheelchair bound by the time of diagnosis and consideration of treatment) have clear initial benefit from decompressive surgery and this benefit is maintained so that there is no later deterioration 3 years following surgery. This is in contrast to unoperated patients who suffer a clear deterioration over the same period. Taking the decision to operate results in initial benefit, and this benefit *increases* over time. Secondly, the study indicates that more severely affected patients may enjoy greater benefit from surgery. Finally, the equal benefit in more elderly patients balanced against a more severe deterioration if untreated, should sway decision-making *towards* operating on elderly patients.

Further advantages of walking data are revealed. Although some of the other scales could reveal the above information, their validity is in doubt because it has not been shown that a change about one baseline level of severity is statistically equivalent to that about a different severity. The continuous numerical data of walking tests means that this assumption can be made and the statistical comparisons need not in general be as conservative. This advantage particularly applies to correlations between measures. For example, in a separate study comparing radiographic with clinical data, instead of simple correlation coefficients, a factorial analysis could be performed. By these means it was

possible to show that the presence of cervical cord signal change on a pre-operative MRI scan could predict a better outcome of surgery than in an otherwise equivalent patient group. Finally, the numerical data allowed a numerical correction to be made for the potential confounding factor of age in each patient, although it was actually found that this had little material effect upon the various findings.

In summary, it is hoped that the demonstrated usefulness of walking data will result in its gaining widespread acceptance in future clinical trials of various interventions in CSM, and perhaps also in other diseases. Its ease of use may mean that with little extra effort, it could also be applied to normal clinical practice in assessing CSM patients, rather than being restricted to a role as a research tool. This would encourage a scientific approach to normal clinical practice, where it would be possible to compare individual patients with trial populations and to gauge operative outcomes against a large background of published research and clinical activity.

Further Studies

The studies described in this thesis lend themselves to a longer term follow-up of a large group of patients. It was clearly not possible in the time span of this work to study a large cohort of patients from progressive recruitment to final follow-up at 3 or perhaps even 5 years. A larger group of patients would enable one to answer questions about the relative benefits of different surgical techniques. In the long-term group described here, out of 60 patients only 50 could be properly analysed and only 7 had laminectomy procedures, not enough for meaningful comparisons. Other important and so far unanswered questions would be the possibility of very late (beyond 3 years) accelerated deterioration as a result of certain procedures or any form of surgical intervention and the role of re-do operations and which re-do operation would be the best to perform. Further investigation of the role that various imaging features might have to play in the prediction of surgical outcome may also be fruitful. While the studies of this thesis have indicated that walking parameters may well be appropriate as stand-alone measures, it might be interesting to investigate the possible advantages of an amalgamation of walking time

with other quantitative parameters such as radiographic measurements, pain analogue scores and hand co-ordination tasks such as the 9-hole peg test. Data from sub-dividing functional measures indicated that the last of these may be a sensitive factor in determining outcome and severity, although perhaps reflecting radiculopathy as well as myelopathy.

Since the first decompressive procedure developed by Victor Horsley in 1892, the 20th Century has seen huge strides made in the development of surgical techniques for treating compressive myelopathy but relatively little in the way of validation of such treatment. It is hoped that, with the introduction of rationalised quantitative assessment such as the walking test, the next 10 years may see such validation finally taking place.

Summary Points

1. Ordinal scales quantifying myelopathic disability have limitations.
2. The walking test, used previously in other conditions, has internal consistency and easily reproducible, it also has sensitivity and specificity. Clinicians want a test that is easy to administer.
3. The walking test shows immediate and long-term benefits from surgery.
4. There is a "group" correlation between severity of MRI scan, walking parameters and surgical outcome.
5. This work suggests that the walking test is a simple, reliable and reproducible test worthy of further, perhaps multi-centre, study.

APPENDIX 1

ASIA Score:

Grade A - complete-no motor function or sensory function in S4 and S5 (bottom segments)

Grade B - sensory but no motor function preserved

Grade C - motor preserved but weak (less than 3/5)

Grade D - motor preserved and greater than or equal to 3/5

Grade E - totally normal

APPENDIX 2

RANAWAT SCALE

I No neurological deficit (normal neurological condition)

II Subjective weakness with hypereflexia and dysesthesia (Parasthesia, Subjective weakness e.g. cannot do buttons up or walk as well as used to, tingling in limbs).

III Objective weakness and long-tract signs but able to walk (on examination definite Weakness and increased tone).

IV Able to walk only with some-one else's help or with aid or frame.

V Chair bound or bedridden.

APPENDIX 3

NURICK'S FUNCTIONAL SCALE

Grade 0: Signs or symptoms of root involvement but without evidence of spinal cord disease

Grade 1: Signs of spinal cord disease but no difficulty in walking.

Grade 2: Slight difficulty in walking, which did not prevent full-time employment

Grade 3: Difficulty in walking which prevented full-time employment or the ability to do all housework, but which is not so severe as to require someone else's help to walk.

Grade 4: Able to walk only with someone else's help or with the aid or frame.

Grade 5: Chair bound or bedridden.

APPENDIX 4

CRITERIA ADOPTED BY THE JAPANESE ORTHOPAEDIC ASSOCIATION FOR EVALUATION OF THE SURGICAL RESULTS FOR CERVICAL MYELOPATHY

DESCRIPTION	GRADE
I Motor dysfunction of the upper extremity	
Unable to feed oneself	0
Unable to handle chop-sticks, able to eat with a spoon	1
Able to handle chopsticks with much difficulty	2
Able to handle chopsticks with slight difficulty	3
None	4
2 Motor dysfunction of the lower extremity	
Unable to walk	0
Can walk on a flat floor with walking aid	1
Can walk up and down/ or down stairs with handrail	2
Lack of stability and smooth gait	3
None	4
3 Sensory deficit	
Upper extremity	
Severe loss of pain	0
Mild sensory loss	1
None	2
Lower extremity	0 - 2
Trunk	0 – 2
4 Sphincter dysfunction	
Unable to void	0
Marked difficulty in micturition (retention)	1
Difficulty in micturition (frequency, hesitancy)	2
None	3

A normal patient scores 17 points.

APPENDIX 5

THE EUROPEAN MYELOPATHY SCORE

Upper motor Neurone function (gait)	
Unable to walk, wheelchair	1
Walking on flat ground only with cane or aid	2
Climbing stairs only with aid	3
Gait clumsy but no aid necessary	4
Normal walking and climbing stairs	5
Upper motor neurone function (bladder and bowel function)	
Retention, no control over bladder and or bowel function	1
Inadequate micturition and urinary frequency	2
Normal bladder and bowel function	3
Lower motor neurone function (hand function)	
Handwriting and eating with knife and fork impossible	1
Handwriting and eating with fork impaired	2
Handwriting, tying shoe laces or a tie clumsy	3
Normal handwriting	4
Posterior column function (proprioception and co-ordination)	
Getting dressed only with aid	1
Getting dressed clumsily and slowly	2
Getting dressed normally	3
Posterior cervical roots (paresthesias and dysesthesia)	
Disabling sensations disturbing all activities	1
Tolerable sensations	2
No paraesthesia or dysesthesia	3
Total	

The maximum score is 18, the minimum 5. Depending on the sum reached in the score, cervical myelopathy is classified.

APPENDIX 6

CLINICAL FOLLOW-UP ODOM'S CRITERIA:

- Excellent results - no complaints referable to cervical disease and were able to carry on their daily occupation without impairment.
- Good results - intermittent discomfort related to cervical disease but did not significantly interfere with their work.
- Satisfactory results - subjective improvement but whose activities were significantly limited.
- Poor results - did not improve or were worse as compared with their condition before operation.

APPENDIX 7

BARTHEL ACTIVITIES OF DAILY LIVING (ADL) INDEX

BOWELS
0 = incontinent (or needs to be given enemata)
1= occasional accident (once a week)
2 = continent

BLADDER
0 = incontinent, or catheterised and unable to manage alone
1 = occasional accident (maximum once per 24 hours)
2 = continent

GROOMING
0 = needs help with personal care
1 = independent face / hair / teeth / shaving (implements provided)

TOILET USE
0 = independent
1 = needs some help, but can do something alone
2 = independent (on and off, dressing, wiping)

FEEDING
0 = unable
1 = needs help cutting, spreading butter, etc

TRANSFER (bed to chair and back)
0 = unable, no sitting balance
1 = major help (one or two people, physical), can sit
2 = minor help (verbal or physical)
3 = independent

MOBILITY
0 = immobile
1 = wheelchair independent, including corners
2 = walks with help of one person (verbal or physical)
3 = independent (but may use any aid; for example, stick)

DRESSING
0 = independent
1 = needs help but can do about half unaided

STAIRS
0 = unable
1 = needs help (verbal, physical, carrying aid)
2 = independent

BATHING
0 = dependent
1 = independent (or in shower)

Total 0 – 20

APPENDIX 8

HEALTH ASSESSMENT QUESTIONNAIRE (HAQ SCORE)

We are interested in learning how your illness effects your ability to function in daily life. Please feel free to add any comments at the end of this form. Please choose from the answers given.

Answers 0 = without *any* difficulty, 1 = with *some* difficulty, 2 = with *much* difficulty, 3 = unable

Item	Answer (score)
1. Dressing and grooming	
Are you able to:	
Dress yourself, including tying shoelaces and doing-up buttons?	
Shampoo your hair?	
2. Rising	
Are you able to:	
Stand up from an armless straight chair?	
Get in and out of bed?	
3. Eating	
Are you able to:	
Cut your meat?	
Lift a full cup or glass to your mouth?	
Open a new carton of milk (or soap powder)?	
4. Walking	
Are you able to:	
Walk outdoors on a flat ground?	
Climb up five steps?	
5. Hygiene	
Are you able to:	
Wash and dry your entire body?	
Take a bath?	
Get on and off the toilet?	
6. Reach	
Are you able to:	
Reach and get down a 5lb (2kg) object from above your head	
(For example, a bag of potatoes)?	
Bend down and pick up clothing from the floor?	
7. Grip	
Are you able to:	
Open car doors?	
Open jars, which have been previously opened?	
Turn taps (faucets) on and off?	
8. Activities	
Are you able to:	
Run errands and go shopping?	
Get in and out of the car?	
Do chores such as vacuuming, housework, or gardening?	

APPENDIX 9

MYELOPATHY DISABILITY INDEX

Please tick the one in response which best describes your usual abilities over the past week	**Without ANY difficulty**	**With SOME difficulty**	**With MUCH difficulty**	**UNABLE to do so**
Score (office use only)	**0**	**1**	**2**	**3**
RISING Are you able to:				
Stand up from an armless straight chair?				
Get in and out of bed?				
EATING Are you able to:				
Cut your meat?				
Lift a full cup or glass to your mouth?				
WALKING Are you able to:				
Walk outdoors on a flat ground				
Climb up five steps?				
HYGIENE Are you able to:				
Wash and dry your entire body?				
Get on and off the toilet?				
GRIP Are you able to:				
Open jars which have been previously been opened?				
ACTIVITIES Are you able to: Get in and out of the car?				
DRESSING Are you able to: Dress yourself, including tying shoelaces, and doing buttons on a shirt or blouse?				
TOTAL (office use only)				

Note: If aids or assistance from another is required to perform any of the tasks please score the activity as "with much difficulty". Total score = A+B+C+D (range 0-33). The final score is expressed as a percentage.

APPENDIX 10

SF 36 HEALTH SURVEY

INSTRUCTIONS: This survey asks for your views about your health. This information will keep track of how you feel and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is: (circle one)

   Excellent 1
   Very good 2
   Good.....3
   Fair.....4
   Poor.....5

2. Compared to year ago, how would you rate your health in general now?

   (circle one)

   Much better than a year ago.....1
   Somewhat better than a year ago.....2
   About the same as a year ago.....3
   Somewhat worse now than a year ago.....4
   Much worse now than a year ago.....5

3. The following items are about the activities you might do during a typical day. Does your health now limit you in these activities? If so how much?

   (circle one number on each line)

ACTIVITIES	**YES Limited a lot**	**YES Limited a little**	**NO not Limited at all**
a. **Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sports	1	2	3
b. **Moderate activities,** such as moving a table, pushing a vacuum cleaner, bowling, or playing golf.	1	2	3
c. Lifting or carrying groceries	1	2	3
d. Climbing **several** flights of stairs	1	2	3
e. Climbing **one** flight of stairs	1	2	3
f. Bending, Kneeling and stooping	1	2	3
g. Walking **more than a mile**	1	2	3
h. Walking **several blocks**	1	2	3
i. Walking **one block**	1	2	3
j. Bathing or dressing yourself	1	2	3

4. During the past month have you had of the following problems with your work or other regular daily activities as a result of your physical health?

(circle one number on each line)

	YES	NO
a. Cut down on the **amount of time** you spent on work and other activities.	1	2
b. **Accomplished** less than you would like	1	2
c. Were limited in the **kind** of work or other activities	1	2
d. Had **difficulty** in performing the work or other activities (for example it took extra effort)	1	2

5. During the past month have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

(circle one number on each line)

	YES	NO
Cut down the **amount of time** you spent on work and other activities	1	2
Accomplished less than you would like	1	2
Didn't do work and other activities as **carefully** as usual	1	2

6. During the past month to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours or groups?

(circle one)

Not at all.....1
Slightly.....2
Moderately.....3
Quite a bit.....4
Extremely.....5

7. How much bodily pain have you had during the past month?

(Circle one)

None.....1
Very mild.....2
Mild.....3
Moderate.....4
Severe.....5
Very Severe.....6

8. During the past month, how much did pain interfere with your normal work (including both work outside the home and housework?)

(Circle one)

Not at all.....1
A little bit.....2
Moderately.....3
Quite a bit.....4
Extremely.....5

9. These questions are about how you feel and how things have been with you during the past month. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past month.

(Circle one number on each line)

	All of the time	**Most of the time**	**A good bit of the time**	**Some of the time**	**A little of the time**	**None of the time**
a. Did you feel full of pep?	1	2	3	4	5	6
b. Have you been a very nervous person?	1	2	3	4	5	6
c. Have you felt down in the dumps that nothing could cheer you up?	1	2	3	4	5	6
d. Have you felt calm and peaceful?	1	2	3	4	5	6
e. Did you have a lot of energy?	1	2	3	4	5	6
f. Have you felt downhearted and blue?	1	2	3	4	5	6
g. Did you feel worn out?	1	2	3	4	5	6
h. Have you been a happy person?	1	2	3	4	5	6
i. Did you feel tired?	1	2	3	4	5	6

10. During the past month, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends and relatives, etc.)?

(Circle one)

All of the time 1
Most of the time.....2
Some of the time.....3
A little of the time.....4
None of the time.....5

11. How TRUE or FALSE is each of the following statements for you?

(Circle one number on each line)

	Definitely True	**Mostly True**	**Don't know**	**Mostly false**	**Definitely True**
a. I seem to get sick a little easier than other people	1	2	3	4	5
b. I am as healthy as anybody I know	1	2	3	4	5
c. I expect my health to get worse	1	2	3	4	5
d. My health is excellent	1	2	3	4	5

APPENDIX 11

Questionnaire (Clinicians assessment of myelopathy):

1. Your Speciality.

2. What in your opinion is the best way to assess function in Cervical Spondylotic Myelopathy patients? Please rank your choice on a 0 to 4 scale (0 being the 'worst' and 4 being the 'best'): History, Clinical Examination, Imaging, Quantitative functional assessment, Unknown/ Don't know.

3. Do you believe there is a "Gold Standard" quantitative scale for assessing Cervical Spondylotic Myelopathy? If YES, which?

4. Do you use an assessment scale(s), if so which one?

5. Please rate the quality/qualities of an assessment scale that make(s) it most useful in your practice. Please rank your choices on 1 to 4 scales. (1 being the worst and 4 being best): Ease of use, Reproducibility, Sensitivity to Change, Validity.

6. Are quantitative scales purely of academic interest, do you actually base clinical decisions on the score, or neither?

7. What other information do you use to make actual management decisions i.e. to perform an operation on the patient or to refer the patient to a surgeon? (Please write your answer below).

APPENDIX 12

Pain: Visual Analogue scale

Reference List

1. (1994) Japanese orthopaedic association scoring system (17-2) for cervical myelopathy. J Jpn Orthop Assoc 490-503

2. al-Mefty O, Harkey HL, Marawi I, Haines DE, Peeler DF, Wilner HI, Smith RR, Holaday HR, Haining JL, Russell WF, et a (1993) Experimental chronic compressive cervical myelopathy. J.Neurosurg. 79: 550-561

3. Alexander JT (1996) Natural history and non-operative management of cervical spondylosis. In: Arnold H.Menezes VKHS (ed) Principles of spinal surgery. McGraw-Hill, New York, pp 547-558

4. Allen K (1952) Neuropathies caused by bony spurs in cervical spine with special reference to surgical treatment. J.Neurol.Neurosurg.Psychiatry 20-36

5. Andrews ET, Gentchos EJ, Beller ML (1971) Results of anterior cervical spine fusions done at the hospital of the University of Pennsylvania. A nine-year follow-up. Clin.Orthop. 8115-20: -20

6. Arnold JG (1955) The clinical manifestation of cervical spondychondrosis (spondylosis) of hte cervical spine. Ann Surg 872-889

7. Baba H, Maezawa Y, Uchida K, Furusawa N, Wada M, Imura S (1997) Plasticity of the spinal cord contributes to neurological improvement after treatment by cervical decompression. A magnetic resonance imaging study. J Neurol 244: 455-460

8. Bailey RW, Bagley CG (1960) Stabilisation of the cervical spine by anterior fusion. J.Bone Joint Surg Am. 565-594

9. Bakay L, Cares HL, Smith RJ (1970) Ossification in the region of the posterior longitudinal ligament as a cause of cervical myelopathy. J.Neurol.Neurosurg.Psychiatry 33: 263-268

10. Ball PA, Saunders RL (1992) The subjective myelopathy. In: Saunders R.L.and Bernini P.M. (ed) Cervical spondylotic myelopathy. Blackwell, Boston, pp 48-55

11. Bednarik J, Kadanka ZVSea (1999) The value of somatosensory and motor evoked potentials in predicting and monitoring the effect of therapy in spondylotic cervical myelopathy. Prospective randomised study. Spine. 1593-1598

12. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1: 307-310

13. Brain WR (1948) Discussion on rupture of the intervertebral disc in the cervical region. Proc R Soc.Med., 509

14. Brain WR, Northfield D, Wilkinson M (1952) The neurological manfestation of cervical spondylosis. Brain 187-225

15. Breig A, Turnbull I, Hassler O (1966) Effects of mechanical stresses on the spinal cord in cervical spondylosis. A study on fresh cadaver material. J.Neurosurg. 25: 45-56

16. Bucciero A, Vizioli L, Tedeschi G (1993) Cord diameters and their significance in prognostication and decisions about management of cervical spondylotic myelopathy. J.Neurosurg.Sci. 37: 223-228

17. Bush JW, Anderson JP, Kaplan RM, Blischke WR (1982) "Counterintuitive" preferences in health-related quality-of-life measurement. Med Care 20: 516-525

18. Campbell AMG, Phillips DG (1960) Cervical disc lesion with neurological disorder. Differential doagnosis, treatment and prognosis. BMJ 481-485

19. Carr RG (1992) The Psychiatrist and cervical spondylosis. In: Saunders R.L.and Bernini P.M. (ed) Cervical spondylotic myeloapthy. Blackwell, Boston, pp 96-109

20. Casey AT, Bland JM, Crockard HA (1996) Development of a functional scoring system for rheumatoid arthritis patients with cervical myelopathy. Ann.Rheum.Dis. 55: 901-906

21. Chiles BW, Leonard MA, Choudhri HF, Cooper PR (1999) Cervical spondylotic myelopathy: patterns of neurological deficit and recovery after anterior cervical decompression. Neurosurgery 44: 762-769

22. Clark CR (1988) Cervical spondylotic myelopathy. Spine 347-349

23. Clarke E, Robinson PK (1956) Cervical myelopathy: a complication of cervical spondylosis. Brain 483-510

24. Clowards RB (1958) The anterior approach for the removal of ruptured cervical discs. J.Neurosurg. 602-617

25. Cronbach LJ, Meehl PE (1955) Construct validity in psychological tests. Psychological Bulletin 281-302

26. Dun-Rankin P (1983) Scaling Methods. Erlbaum, Hilldale, New Jersey

27. Dunsker SB (1981) Cervical spondylotic myelopathy: pathogenesis and pathophysiology. Raven Press, New York

28. Emery SE, Bohlman HH, Bolesta MJ, Jones PK (1998) Anterior cervical decompression and arthrodesis for the treatment of cervical spondylotic myelopathy. Two to seventeen-year follow-up. J.Bone Joint Surg Am. 80: 941-951

29. Epstein JA, Carras R, Lavine LS, Epstein BS (1969) The importance of removing osteophytes as part of the surgical treatment of myeloradiculopathy in cervical spondylosis. J.Neurosurg. 30: 219-226

30. Fager CA (1973) Results of adequate posterior decompression in the relief of spondylotic cervical myelopathy. J.Neurosurg. 38: 684-692

31. Fearnside MR (2000) Decompressive cervical laminectomy. In: Kaye AH and Black PMcL (ed) Operative Neurosurgery. Harcourt Publishers Ltd, London, pp 1773-1784

32. Freidenberg ZB, Miller WT (1963) 'Degenerative disease of the cervical spine'. J Bone Joint Surg. 1171-1178

33. Froberg DG, Kane RL (1989) Methodology for measuring health-state preferences--I: Measurement strategies. J.Clin.Epidemiol. 42: 345-354

34. Fujiwara K, Yonenobu K, Ebara S, Yamashita K, Ono K (1989) The prognosis of surgery for cervical compression myelopathy. An analysis of the factors involved. J.Bone Joint Surg.Br. 71: 393-398

35. Fukushima T, Ikata T, Taoka Y, Takata S (1991) Magnetic resonance imaging study on spinal cord plasticity in patients with cervical compression myelopathy. Spine. 16: S534-S538

36. Goldfarb BJ, Simon SR (1984) Gait patterns in patients with amyotrophic lateral sclerosis. Arch.Phys.Med.Rehabil. 65: 61-65

37. Gonzales-Feria L (1975) The effect of surgical immobilisation after laminectomy in treatment of advanced cases of cervical spondylotic myelopathy. Acta Neurochir 185-193

38. Gooding MR, Wilson C, Hoff J (1972) Experimental cervical myelopathy: Autogradiographic stuides of spinal cord blood flow patterns. Surg Neurol 233-239

39. Gooding MR, Wilson CB, Hoff JT (1975) Experimental cervical myelopathy. Effects of ischemia and compression of the canine cervical spinal cord. J.Neurosurg. 43: 9-17

40. Guyatt G, Walter S, Norman G (1987) Measuring change over time: assessing the usefulness of evaluative instruments. J.Chronic.Dis. 40: 171-178

41. Hadley MN, Sonntag KH (1996) Cervical Disc Herniation: The Anterior Approach without Interbody fusion. In: Arnold H Menezes and Volker K H Sonntag (ed) Principles of Spinal Surgery. McGraw-Hill, New York, pp 531-538

42. Hamburger C, Buttner A, Uhl E (1997) The cross-sectional area of the cervical spinal canal in patients with cervical spondylotic myelopathy. Correlation of preoperative and postoperative area with clinical symptoms. Spine. 22: 1990-1994

43. Harsh GR, Sypert GW, Weinstein PR, Ross DA, Wilson CB (1987) Cervical spine stenosis secondary to ossification of the posterior longitudinal ligament. J.Neurosurg. 67: 349-357

44. Hashizume Y, Iijima S, Kishimoto H, Yanagi T (1984) Pathology of spinal cord lesions caused by ossification of the posterior longitudinal ligament. Acta Neuropathol.(Berl) 63: 123-130

45. Herdman J, Linzbach M, Krzan M, et al (1994) The European Myelopathy score. In: Bauer BBMKM (ed) Advances in neurosurgery. Springer, Berlin, pp 266-268

46. Hirabayashi K, Miyakawa J, Satomi K, Maruyama T, Wakano K (1981) Operative results and postoperative progression of ossification among patients with ossification of cervical posterior longitudinal ligament. Spine. 6: 354-364

47. Hirabayashi K, Toyama Y, Chiba K (1999) Expansive laminoplasty for myelopathy in ossification of the longitudinal ligament. Clin.Orthop. 35-48

48. Hirabayashi K, Watanabe K, Wakano K, Suzuki N, Satomi K, Ishii Y (1983) Expansive open-door laminoplasty for cervical spinal stenotic myelopathy. Spine. 693-699

49. Hitselberger WE, Witten RM (1968) Abnormal myelograms in asymptomatic patients. J.Neurosurg. 28: 204-206

50. Holden MK, Gill KM, Magliozzi MR, Nathan J, Piehl BL (1984) Clinical gait assessment in the neurologically impaired. Reliability and meaningfulness. Phys.Ther. 64: 35-40

51. Horseley V (1892) Historical introduction. In: Brain W.R WM (ed) Cervical Spondylosis. Heinemann, London, pp 1-9

52. Hughes JTBB (1965) Necropsy observations on the spinal cord in cervical spondylosis. Riv Patol Nerv Ment 196-204

53. Hukuda S, Wilson CB (1972) Experimental cervical myelopathy: effects of compression and ischemia on the canine cervical cord. J.Neurosurg. 37: 631-652

54. Hunt WE (1980) Cervical spondylosis: natural history and rare indications for surgical decompression. Clin.Neurosurg. 27: 466-480

55. Huskisson EC (1974) Measurement of pain. Lancet 2: 1127-1131

56. Irvine HDFJBNDJea (1965) Prevalence of cervical spondylosis in a general practice. Lancet 1089-92

57. Kadanka ZBJVSeal (2000) Conservative treatment versus surgery in spondylotic cervical myelopathy: a prospective randomised study. Eur Spine J. 538-544

58. Kahn EA (1947) The role of dentate ligaments in spinal cord compression and the syndrome of lateral screlosis. J Neurosurg 191-199

59. Kaiser JA, Holland BA (1998) Imaging of the cervical spine. Spine. 23: 2701-2712

60. Key C.A. On paraplegia depending on disease of the spine. 3, 17-34. 1838. Guys Hospital Report.
Ref Type: Report

61. Kuhtz BJP, Johnk K, Mader S, Stolze H, Mehdorn M (1999) Analysis of gait in cervical myelopathy. Gait.Posture. 9: 184-189

62. Kumar VG, Rea GL, Mervis LJ, McGregor JM (1999) Cervical spondylotic myelopathy: functional and radiographic long-term outcome after laminectomy and posterior fusion. Neurosurgery 44: 771-777

63. La Rocca NG (1989) Statistical and methodologic considerations in scale construction. In: Munsat TL (ed) In Quantification of Neurologic Deficit. Butterworths, Stoneham, pp 49-67

64. Lawrence JS, Bremner JM, Bier F (1966) 'Osteo-arthrosis: Prevalence in teh population and relationship between symptoms and x-ray changes'. Ann.Rheum.Dis. 1-24

65. Lees F, Turner JWA (1963) Natural history and prognosis of cervical spondylosis. BMJ 1607-1610

66. Levine DN (1997) Pathogenesis of cervical spondylotic myelopathy. J Neurol.Neurosurg Psychiatry 62: 334-340

67. Lindsay KW, Bone I, Callander R (1991) Disc prolapse and spondylosis. In: Lindsay K.W. BIaCR (ed) Neurology and Neurosurgery Illustrated. Churchill Livingstone, Edinburgh, pp 389-398

68. Lunsford LD, Bissonette DJ, Jannetta PJ, Sheptak PE, Zorub DS (1980b) Anterior surgery for cervical disc disease. Part 1: Treatment of lateral cervical disc herniation in 253 cases. J.Neurosurg. 53: 1-11

69. Lunsford LD, Bissonette DJ, Zorub DS (1980a) Anterior surgery for cervical disc disease. Part 2: Treatment of cervical spondylotic myelopathy in 32 cases. J.Neurosurg. 53: 12-19

70. MacKenzie CR, Charlson ME (1986) Standards for the use of ordinal scales in clinical trials. Br.Med.J.Clin.Res.Ed. 292: 40-43

71. Martin J, Leltzer, Elliot D (1988) The Prevalence of disability among Adults: Office of population censuses and surveys. HMSO, London

72. May D, Nayak US, Isaacs B (1985) The life-space diary: a measure of mobility in old people at home. Int.Rehabil.Med 7: 182-186

73. Montgomery DM, Brower RS (1992) Cervical spondylotic myelopathy. Clinical syndrome and natural history. Orthop.Clin.North Am. 23: 487-493

74. Naderi S, Ozgen S, Pamir MN, Ozek MM, Erzen C (1998) Cervical spondylotic myelopathy: surgical results and factors affecting prognosis. Neurosurgery 43: 43-49

75. Nurick S (1972) The pathogenesis of the spinal cord disorder associated with cervical spondylosis. Brain 95: 87-100

76. Odom GL, Finney W, Woodhall B (1958) Cervical disc lesions. Journal of American Medical Association 879-884

77. Ogino H, Tada K, Okada K, Yonenobu K, Yamamoto T, Ono K, Namiki H (1983) Canal diameter, anteroposterior compression ratio, and spondylotic myelopathy of the cervical spine. Spine. 8: 1-15

78. Ohtsuka K, Terayama K, Yanagihara M, Wada K, Kasuga K, Machida T, Furukawa K (1986) An epidemiological survey on ossification of ligaments in the cervical and thoracic spine in individuals over 50 years of age. Nippon Seikeigeka Gakkai Zasshi 60: 1087-1098

79. Okada Y, Ikata T, Yamada H, Sakamoto R, Katoh S (1993) Magnetic resonance imaging study on the results of surgery for cervical compression myelopathy. Spine. 18: 2024-2029

80. Pallis C JASJ (1954) Cervical Spondylosis. Brain 274

81. Payne EE, Spillane JD (1957) The cervical spine: An antomicopathological study of the 70 specimens (using a special technique) with particular reference to the problem of cervical spondylosis. Brain 571-596

82. Persson LCGCCACJA (1997) Long lasting cervical radicular pain managed with surgery, physiotherapy or a cervical collar. Spine 751-8

83. Phillips DG (1973) Surgical treatment of myelopathy with cervical spondylosis. J.Neurol.Neurosurg.Psychiatry 36: 879-884

84. Piepgras DG (1977) Posterior decompression for myelopathy due to cervical spondylosis: laminectomy alone versus laminectomy with dentate ligament section. Clin.Neurosurg. 24508-15: -15

85. Ranawat COPPPeal (1979) Cervical fusion in rheumatoid arthritis. Journal of Bone and Joint Surgery America 1003-1110

86. Rengachary, S. S. and Redford, J. B. Partial median corpectomy for cervical spondylotic myelopathy. (In Neurosurgery Update II). 1991. New York, McGraw-Hill.
Ref Type: Generic

87. Robinson RA, Smith GW (1955) Abtero-lateral cervical disc removal and interbody fusion for cervical disc syndrome. John Hopkins 223-224

88. Robinson RA, Walker AE, Ferlic DCeal (1962) The results of anterior interbody fusion of the cervical spine. J Bone Joint Surg Am. 1569-1587

89. Rowland LP (1992) Surgical treatment of cervical spondylotic myelopathy: time for a controlled trial. Neurology 42: 5-13

90. Sampath P, Bendebba M, Davis JD, Ducker T (1999) Outcome in patients with cervical radiculopathy. Prospective, multicenter study with independent clinical review. Spine. 24: 591-597

91. Singh A, Crockard HA (1999) Quantitative assessment of cervical spondylotic myelopathy by a simple walking test. Lancet 354: 370-373

92. Singh A, Crockard HA (2001) Comparison of seven different scales used to quatify severity of cervical spondylotic myelopathy and post-operative improvement. Journal of outcome measurement 798-818

93. Singh ACHAPASJ (2001) Related ArticlesClinical and radiological correlates of severity and surgery-related outcome in cervical spondylosis. J Neurosurg. 94: 189-198

94. Steinbroker O, TRC, Batterman RC (1949) Therapeutic criteria in rheumatoid arthritis. Journal of American Medical Association 659-662

95. Steiner DI, Norman GR (1996) Measuring change in health measurement scales: a practical guide to their development and use. Oxford University Press, Oxford

96. Stoltmann HFBW (1964) The role of the ligamentuaflava in the pathogenesis of myelopathy in cervical spondylosis. Brain 45

97. Stookey B (1928) Compression of the spinal cord due to ventral extradural cervical chondromas: Diagnosis and surgical treatment. Arch Neurol Psychiatry 275-291

98. Symon L, Lavender P (1967) The surgical treatment of cervical spondylotic myelopathy. Neurology 17: 117-127

99. Sypert GW (1992) Anterior decompression and fusion for cervical myelopathy. In: Camins MB OP (ed) Disorders of the cervical spine. Williams & Wilkins, Baltimore, pp 407-416

100. Sypert GW, Arpin-Sypert EJ (1996) Indications of corpectomy in the cervical spine. In: Arnold H.Menezes VKHS (ed) Principles of spinal surgery. McGraw-Hill, New York, pp 1147-1166

101. Takahashi M, Sakamoto Y, Miyawaki M, Bussaka H (1987) Increased MR signal intensity secondary to chronic cervical cord compression. Neuroradiology. 29: 550-556

102. Teresi LM, Lufkin RB, Reicher MA, Moffit BJ, Vinuela FV, Wilson GM, Bentson JR, Hanafee WN (1987) Asymptomatic degenerative disk disease and spondylosis of the cervical spine: MR imaging. Radiology 164: 83-88

103. Tomassi P (1995) Logic and scientific method. In: Phillips CI (ed) Logic in medicine. BMJ Publishing Group, London, pp 31-58

104. Travell J, Simons DG (1983) Myofacial pain dysfunction, the trigger point manual. Williams and Wilkins, Baltimore

105. Wada E, Yonenobu K, Suzuki S, Kanazawa A, Ochi T (1999) Can intramedullary signal change on magnetic resonance imaging predict surgical outcome in cervical spondylotic myelopathy? Spine. 24: 455-461

106. Wade DT (1992) Measurement in neurological rehabilitation. Oxford Unviersity Press, Oxford

107. Wade DT, Wood VA, Heller A, Maggs J, Langton HR (1987) Walking after stroke. Measurement and recovery over the first 3 months. Scand.J.Rehabil.Med. 19: 25-30

108. Ware JEJr, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med.Care 30: 473-483

109. Wassertheil-Smoller S (1995) Biostatics nad Epidemiology - A Primer for Health Professionals. Springer-Verlag, New York

110. Yonenobu K, Fuji T, Ono K, Okada K, Yamamoto T, Harada N (1985) Choice of surgical treatment for multisegmental cervical spondylotic myelopathy. Spine 10: 710-716

111. Yu YL, du BG, Stevens JM, Kendall BE (1986) Computed tomography in cervical spondylotic myelopathy and radiculopathy: visualisation of structures, myelographic comparison, cord measurements and clinical utility. Neuroradiology. 28: 221-236

112. Zeidman SM, Ducker TM (1992) Cervical Disc diseases: Part I. Treatment options and outcomes. Neurosurg Q. 116-143