The pathophysiology of intervertebral disc degeneration has been extensively studied. Various factors have been suggested as influencing its aetiology, including mechanical factors, such as compressive loading, shear stress and vibration, as well as ageing, genetic, systemic and toxic factors, which can lead to degeneration of the disc through biochemical reactions. How are these factors linked? What is their individual importance? There is no clear evidence indicating whether ageing in the presence of repetitive injury or repetitive injury in the absence of ageing plays a greater role in the degenerative process. Mechanical factors can trigger biochemical reactions which, in turn, may promote the normal biological changes of ageing, which can also be accelerated by genetic factors. Degradation of the molecular structure of the disc during ageing renders it more susceptible to superimposed mechanical injuries.

This review supports the theory that degeneration of the disc has a complex multifactorial aetiology. Which factors initiate the events in the degenerative cascade is a question that remains unanswered, but most evidence points to an age-related process influenced primarily by mechanical and genetic factors.

Disc degeneration is common, but a universally accepted definition has proved elusive. For surgeons and radiologists, degeneration might mean the presence of osteophytes and loss of signal intensity on MRI. To a biochemist, it may be expressed by changes in the content of proteoglycans or water. To a pathologist, the disc is dry, with cracks and fissures. The reason for this variance is that different disciplines use different tools, and hence see different things. This review attempts to provide a unified model of disc degeneration, bringing together all the known facts and accommodating many of the conjectures about its nature. It attempts to redress some of the misconceptions. To what degree the process of disc degeneration can contribute to back pain and disc herniation are outside the scope of this paper.

A thorough review of the literature up to January 2008 was carried out using electronic and manual searches. We retrieved and reviewed 3212 papers. Only 169 original articles addressing the pathology of disc degeneration were deemed pertinent.

The normal disc

At a molecular and cellular level, the lumbar intervertebral disc has similar constituents to articular cartilage. Chondrocyte-like cells synthesise type II collagen, proteoglycans, and non-collagenous proteins that collectively form the matrix of the nucleus pulposus and the cartilaginous vertebral endplate. Fibroblast-like cells synthesise type I and type II collagen for the annulus fibrosus (Fig. 1). Proteoglycans consist of a core protein from which radiate chains of glycosaminoglycans containing keratan sulphate and chondroitin sulphate. Multiple proteoglycans are joined to a hyaluronic acid chain to form aggregates. Aggregates are held together by type II collagen, which is cross-linked by type IX collagen.1 The hydroscopic properties of the proteoglycan matrix endow the nucleus with hydrostatic properties,2 allowing it to accommodate compression loads and to brace the annulus. However, the constituents of the matrix are not static. They are continually degraded by enzymes, the matrix metalloproteinases (MMPs), which are secreted by the chondrocytes.3,4 Degradation of the matrix allows it to be refreshed by newly-synthesised components.

Growth factors, such as basic fibroblast growth factor (bFGF), transforming growth factor (TGF) and insulin-like growth factor (IGF), stimulate the chondrocyte or fibroblast...
to produce more matrix, and inhibit the production of MMPs. These growth factors are normally bound by cartilage intermediate layer protein (CLIP) and are released if the matrix is degraded, in order to promote further synthesis. Tissue inhibitors of metalloproteinases (TIMP) suppress the activation of MMPs, thereby controlling degradation. Decreases in pH lessen the rate of synthesis of matrix proteoglycans.

Cytokines, such as interleukin-1 (IL-1), interferon (IFN), and tumour necrosis factor-α (TNF-α) inhibit the synthesis of the matrix and promote the production of MMPs. These cytokines are produced by macrophages which enter the disc.
in response to injury.9 Macrophages also secrete superoxide \( \text{O}_2^- \), which can degrade hyaluronic acid and proteoglycans, causing them to deaggregate, and can inhibit chondrocyte proliferation and synthesis. TNF-\( \alpha \) and IL-1 stimulate inducible nitric oxide synthetase to produce nitric oxide,10 which has a variety of degradative effects. It affects matrix constituents directly, inhibits TIMPs, and thereby promotes matrix degradation and inhibits matrix synthesis.10,11

**Degeneration**

Disc degeneration will occur if the matrix is not normal. This can arise if the components synthesised are themselves abnormal, or if the balance between synthesis and degradation of normal components is disturbed in favour of degradation. At a molecular level, degeneration will be expressed by the production of abnormal components of the matrix or by an increase in the mediators of matrix degradation (IL-1, TNF-\( \alpha \), superoxide and nitric oxide)10 and of MMPs and a reduction in the levels of TIMPs.12,13 Meanwhile, increased concentrations of growth factors (bFGF, TGF) will reflect attempts to repair the degraded matrix.14,15 At macroscopic and microscopic levels degeneration will become evident in the form of structural changes and defects caused by the altered matrix or impaired function of the disc (Fig. 1).

Several factors have been implicated or postulated as causing disc degeneration.

**Ageing.** Changes occur as the disc ages. Individually and collectively these changes reflect impaired synthesis of the matrix. The concentration of cells in the disc declines with age, especially in the annulus.16-18 The rate of synthesis of proteoglycans decreases,19 as does the concentration of proteoglycans in the nucleus.20-22 The proteoglycans produced are smaller24-26 and less aggregated27 because of a decline in link proteins21 and type IX collagen.27,28 The concentration of chondroitin sulphate falls, resulting in a rise in the ratio of keratan sulphate to chondroitin sulphate.22,29,30

The collagen content of the nucleus increases and changes from type II to type I,21,31-34 rendering the nucleus more fibrous.35 The distinction between the nucleus and annulus becomes less apparent as the two regions coalesce.36 Non-collagenous proteins in the nucleus increase.37-41 Increased collagen and increased collagen-proteoglycan binding leave fewer polar groups of the proteoglycans available to bind water.31 The nucleus becomes progressively more solid, dry and granular,36 and cracks appear in the desiccated, fibrous nucleus.35 The collagen lamellae of the annulus increase in thickness and become increasingly fibrillated.32,45 Cracks and cavities may develop within it.21,45

Although these changes have been extensively described, their cause is not known. Among the possibilities that have been raised are declining nutrition, cell senescence, accumulation of degraded matrix products and fatigue failure of the nucleus.21

Cells from the intervertebral discs are subject to senescence and lose their ability to proliferate.46,47 Senescent cells may induce degeneration by decreased anabolism or increased catabolism.47-49 Herniated discs are associated with an increased degree of senescence.47 The accumulated senescent cells reduce the ability of the disc to replace those lost to necrosis or apoptosis.46

However, some features of disc degeneration are not age-related. Narrowing of the intervertebral discs has previously been considered one of the signs of ageing of the lumbar spine,50,51 but post-mortem studies have shown that lumbar discs do not narrow with age,52 implying that a process other than ageing is responsible. Similarly, although pathologists regard tears of the annulus as a degenerative change,35,36 it has been shown that radial tears do not correlate with age.52 Such fissures are indicative of another process. Furthermore, although all discs are the same age, those at lower lumbar levels exhibit degenerative changes far more often than in the upper levels.51 This indicates mechanical loading as the causative factor, rather than simply ageing.

**Genetic factors.** Studies in twins have shown a genetic predisposition to disc degeneration.53-58 Environmental factors only have a modest effect in identical twins.55 These population studies have prompted searches for how genetic factors might cause degeneration. A few possible mechanisms have been identified, particularly genes encoding molecules related to the properties of the extracellular matrix. Taq I and Fok I of the vitamin D receptor gene have been implicated in disc degeneration in several studies.53,59,60 Taq I increases the decay by 30% of the messenger RNA that signals vitamin D efficiency,61 which impairs the sulphation of glycosaminoglycans by vitamin D.62 The prevalence of the t-allele of Taq I, and hence the risk for disc degeneration, differs among races. It is present in 43% of Caucasians and 31% of Africans, but in only 8% of Asians.63

Polymorphism (5A and 6A alleles) commonly occurs in the promoter region of the gene that regulates MMP-3 production.64 The 5A allele is a possible risk factor for accelerated degenerative changes of lumbar discs in the elderly, but not in the young.65

Trp2 and Trp3 alleles result in the substitution by tryptophan of the amino acids in type IX collagen encoded by COL9A2 and COL9A3 genes. Patients with these variants of type IX collagen have an increased risk for lumbar disc disease and chronic sciatica.66,67 The tryptophan-containing type IX collagen gene probably produces an unstable triple helix that increases the susceptibility of disc tissue to mechanical stress.68 Those who carry the Trp3 allele are at an increased risk of developing disc degeneration if they are obese,69 implicating an interplay between genes and environmental factors. The prevalence of Trp alleles and their relevance as risk factors varies with ethnicity. Trp3 occurs in 24% of Finnish patients with disc degeneration,66 but is absent in the southern Chinese.70 Conversely, Trp2 occurs in 20% of southern Chinese
patients with disc degeneration, but in only 4% of Finnish patients suffering from sciatica. Meanwhile, studies in Greek populations have refuted Trp2 and Trp3 alleles as major risk factors for disc degeneration.

Chondroitin sulphate (CS) chains are present in two adjacent regions of the proteoglycan molecule: the CS1 and CS2 domains. In the human AGCl gene, the region coding for the CS1 domain exhibits specific polymorphism which is associated with disc degeneration. The prevalence of this gene in populations has not been determined.

A single nucleotide polymorphism (1184C allele) results in a amino acid substitution in the CLIP gene that encodes cartilage intermediate layer protein. The presence of an 1184C allele increases the binding and inhibition of growth factors such as TGF-β, reducing the growth factor-mediated induction of matrix formation genes.

Among the genes that code for interleukin-1, the alleles (IL-1αT889 and IL-1βT3954) are associated with disc bulging, with odds ratios of 2.4 and 3.0, respectively, suggesting a genetic predisposition to disc degeneration through alterations in the function of pro-inflammatory mediators.

Nutrition. The intervertebral disc is the largest avascular tissue in the body. Cells in the centre of an adult lumbar disc are approximately 8 mm away from the nearest blood supply. Cells in the outer annulus obtain nutrients from blood vessels in the soft tissues around its periphery, and from a sparse penetration of capillaries into its outermost region. The nucleus and the cells of the inner annulus rely on a more complicated path extending from the blood vessels of the vertebral body to a capillary network that penetrates the subchondral plate. Nutrients diffuse from these capillaries across the cartilaginous endplate and into the disc. One reason for degeneration is a reduction in the transport of nutrients into the disc.

Epidemiological and post-mortem angiographic studies indicate that insufficient blood supply to the lumbar spine due to atheromatous lesions in the abdominal aorta, or congenital hypoplasia of the lumbar arteries, could be a causative factor in disc degeneration. The capillary network at the bone-cartilage endplate junction diminishes after the first decade of life when the first signs of disc degeneration become evident. Calcification of the endplate occludes the vascular openings within it, acting as a barrier to transport of nutrients because of lowering of endplate permeability. Calcification of the endplate in scoliotic patients has been correlated with loss of nutrient and cell death. The capillary network is regulated by vasoactive agents such as noradrenaline (norepinephrine) and acetylcholine, and by mechanical stimuli such as vibration. Smoking is associated with an increased incidence of disc herniation, implying a systemic effect by constriction of arterioles or anoxia to cells in the disc induced by carboxyhaemoglobin. In animal models, remodelling of this capillary bed has been demonstrated in response to nicotine.

The disc relies on diffusion for its nutrition, and pivotal to this is movement, which pumps water and nutrients into the disc. Accordingly, some investigators have considered that sustained compression, or immobilisation even without compression, might be the basis of impaired nutrition to the disc. However, the laboratory data are conflicting. Studies in rabbits and mice have found that constant loading in compression resulted in disc degeneration, but a study in dogs found that high, static compressive forces across lumbar discs for up to a year did not produce annular fissures, bulging or any other visible form of disc degeneration. It has also been shown that there is reduced synthesis of proteoglycans and an increase in MMP-3 after the application of pressures above and below the physiological levels. These data are of limited clinical relevance because human subjects are unlikely to suffer prolonged static loads.

A reduced supply of nutrients leads to an increase in oxidative stress markers, seen initially in the nucleus of young discs, and eventually throughout the inner regions of the disc with increased age and degeneration. Deposition of carboxymethyllysine reflects a distinct cellular reaction to increased oxidative stress, and possibly impaired function. A low level of oxygen and an acidic pH from the anaerobic metabolism lead to a fall in protein and synthesis of proteoglycan. A fall in nutrient supply can also reduce the number of viable cells in the disc.

Toxic factors. Nicotine directly inhibits proliferation of disc cells and their synthesis of extracellular matrix, as shown on bovine cells from the nucleus pulposus cultured in vitro. Similarly, passive smoking in rats resulted in downregulation of collagen genes which preceded the histological changes of degeneration.

Metabolic disorders. Various metabolic disorders can cause disc degeneration either by interfering with the normal biochemistry of matrix synthesis or by deposition of foreign materials in the disc.

In patients with diabetes mellitus, the nucleus pulposus demonstrates a significant decrease in hexosamine content, an increase in hydroxyproline and enhanced activity of enzymes involved in the metabolism of carbohydrates. The discs exhibit deficiencies in incorporation of 35S-sulphate during proteoglycan synthesis, which indicates a reduced rate of glycosylation and a decrease in the number of sugar side chains per core protein.

Patients with alkaptonuria develop intradiscal deposits of a black pigment, inducing disc degeneration characterised by calcification appearing as elliptical opaque wafers.

Low-grade infection. It has been proposed that disc degeneration could be caused by low-grade infection. One group found that 31% of discs harvested after microdiscectomy tested positive for low-virulence Gram-positive bacteria, and 84% of these were infected with Propionibacterium
acnes. Fritzell et al. could not confirm these findings, although two of the ten patients they studied exhibited evidence of bacterial infection unrelated to P. acnes.

Patients who recover poorly after surgery for disc herniation have exhibited elevated serum concentrations of high-sensitivity C-reactive protein (hs-CRP). This could be caused by interleukin-6, which occurs in degenerative discs. However, it is not clear whether CRP is a sign of disc infection or an inflammatory response to disc material in the epidural space.

**Neurogenic inflammation.** Upon mechanical stimulation, cells in the dorsal root ganglion produce substance P, which acts centrally as a neurotransmitter. However, some investigators have proposed that antidiromic release of substance P into the disc innervated by the ganglion might produce degeneration by stimulating the synthesis of inflammatory agents and degradative enzymes in the disc. This phenomenon is unlikely to be a primary cause of degeneration because it requires a cause of irritation of the ganglion before the disc degenerates or herniates, but disc herniation is the most common cause of this. Nevertheless, it is possible that neurogenic inflammation might aggravate disc degeneration once a herniation or osteophyte irritates the ganglion.

**Autoimmune theory.** Damage to the endplate after a compression injury exposes the antigenic components of the nucleus pulposus to the circulation, which may trigger an immune response. This may be responsible for a chronic inflammatory reaction leading to resorption of the affected disc, or may even involve discs at different vertebral levels, resulting in multilevel degeneration. Local expression of Fas ligand by cells in the disc plays a key role in the molecular mechanism that maintains the immune-privileged characteristics of the disc by inducing apoptosis of invading Fas-positive T cells. Animal experiments have confirmed that violation of the physiological barrier of the disc by mechanical injury changes the role of Fas ligand and induces apoptosis of the disc cells.

**Mechanical factors.** Circumstantial evidence suggests the possibility of mechanical factors in the aetiology of disc degeneration. Degeneration is more common and more severe at lower lumbar levels. Discs immediately above a transitional vertebra tend to exhibit significantly more degenerative changes than those between the transitional vertebra and the sacrum. Mechanical factors are postulated in the production of disc degeneration adjacent to a lumbar fusion.

**Vibration.** Vibration has been incriminated in the pathogenesis of disc degeneration. Results of animal and in vitro studies suggest that vibration can adversely affect the nutrition and metabolism of the disc, especially if the vibration matches the resonant frequency of the lumbar spine (4 Hz to 6 Hz). People exposed to whole-body vibration in the resonant range, such as helicopter pilots and drivers of trucks, buses and tractors, have a high rate of back pain. Similarly, epidemiological studies have provided some support for the hypothesis that driving adversely affects the intervertebral discs, with a higher rate of herniation in occupational drivers. However, MRI observation in identical twins with different patterns of occupational driving do not support the hypothesis that driving is a risk factor for disc degeneration. Epidemiological studies on tractor-driving farmers, rally drivers, and operators of heavy earth-moving machinery have not revealed an association between long-term exposure to vibration and disc degeneration. Back pain related to vibration exposure may be related to other factors.

**Torsion.** Torsional movements generate tension in half of the collagen fibres in the annulus, whereas the other fibres tend to become slack. For the fibres to incur damage, they must be elongated by more than 4% of their resting length, and this is feasible only when axial rotation at each individual segment exceeds 3°. For this to occur, the zygapophyseal joints must be damaged first, as they resist rotation beyond 1.5° to 3°. However, when the spine is flexed, the gaping facet joints offer less constraint to rotation, leading to an annular tear without damage to the facets.

In cadaver studies torsional injury resulted in tears in the periphery of the annulus, usually in the posterolateral region, which slowly extended radially towards the centre of the disc with further torsion. In the degenerative spine, the tear may start in the inner fibres of the annulus and extend toward the periphery. In animal studies, rotation of the lumbar spine after facetectomy leads to degenerative changes, with narrowing of the disc space, fissuring of the annulus and migration of the nucleus.

That torsion injury may lead to disc degeneration can be supported by clinical imaging studies. Professional bowlers in cricket with a chronic stress reaction of the pars interarticularis or a unilateral stress fracture tend to have an intervertebral disc of normal height and appearance on MRI. However, bilateral stress fractures are associated with severe disc degeneration. As cadaver studies have shown that the apophyseal joints resist most of the intervertebral shear, it can be assumed that increased shear stress, presumably torsional injuries caused by unilateral pars fracture, can precipitate disc degeneration.

**Compression.** The lumbar disc is designed to sustain compression loads which are beneficial to the disc. Loading is the physiological stimulus for matrix turnover. It induces MMPs and nitric oxide, ostensibly to clear the matrix in preparation for new synthesis, and induces matrix synthesis. However, excessive loading can lead to deleterious changes in the disc by reducing gene expression of all anabolic proteins with significant effects on aggrecan formation, while simultaneously increasing gene expression of MMPs.

The most vulnerable component of a lumbar disc is the vertebral endplate. When subjected to compression, the endplate fails by fracturing. This can occur as a result of...
application of a sudden, severe compression load\textsuperscript{156-158} or as a result of fatigue failure.\textsuperscript{158,159} Repeated application of loads amounting to between 50\% and 80\% of the ultimate tensile strength of the endplate can cause a fracture after as few as 100 cycles.\textsuperscript{160} Subsequently, endplate fractures can precipitate degeneration by a variety of mechanisms. Callus formation might occlude blood vessels in the endplate, thereby interfering with cell nutrition and maintenance of the extracellular matrix. An \textit{in vitro} study showed that damage to the endplate can produce adverse effects on diffusion into the disc; morphological features of degeneration correlated inversely with the density of vascular openings in the endplate.\textsuperscript{87} Studies both in humans\textsuperscript{161} and in pigs\textsuperscript{162} have related the severity of disc degeneration to that of the injury to the endplate. There is a strong association between degeneration and defects in the endplate from Schmorl’s nodes,\textsuperscript{163-165} Scheuermann’s disease,\textsuperscript{166} and fractures,\textsuperscript{167} with an increased incidence of disc prolapse, particularly at the lower lumbar levels.\textsuperscript{164}

Damage to the endplate rapidly leads to depressurisation of the nucleus and a simultaneous increase in stress in the posterior annulus.\textsuperscript{158,159} The depressurised nucleus is no longer able to brace the annulus, as a result of which the inner lamellae of the annulus buckle inwards and the outer lamellae buckle outwards, particularly posteriorly. Finite element models have confirmed that interlaminar shear stresses arising from a compressive load are highest in the posterolateral annulus.\textsuperscript{168} These stresses lead to separation of the adjacent laminae (delamination), which appears as concentric tears in the annulus fibrosus.\textsuperscript{163}

Mechanisms have been postulated whereby fractures of the endplate can lead to biochemical changes in the matrix of the disc. Mechanobiological studies both \textit{in vivo} and \textit{in vitro} have clearly demonstrated that compression can influence the biosynthetic activity of cells in the disc, altering the expression of key extracellular matrix molecules.\textsuperscript{99,169,170} Furthermore, destruction of cartilage from an endplate fracture would provoke an IL-1-mediated inflammatory response, inducing enzymes that destroy proteoglycans.\textsuperscript{9} If this response is extended into the adjacent matrix, it could initiate degradation.\textsuperscript{118,144} Another possibility is that exposure of the matrix to the blood in the vertebral body might
Degeneration is not a diagnosis but an expression of the state of the disc, which is the result of several factors acting individually or collectively (Fig. 2). Rather than being the result of a single process, disc degeneration can have a number of possible causes. Proper diagnosis requires that the exact aetiology is clearly established.

Degeneration secondary to metabolic disorders does not pose a diagnostic problem as the general features of the condition provide the diagnosis. Where the difficulty arises is in determining whether the condition is genetic, due simply to ageing, or traumatic.

Disc degeneration can be due to genetic factors that produce abnormal components of the matrix which compromise the structure and function of the disc. However, degeneration cannot be attributed wholly to genetic factors. Epidemiological studies have shown that genetic factors increase the risk of degeneration, but they do not account for all cases, nor are they uniform across different ethnic populations. Studies across large multiethnic populations are required.

The classic interpretation has been that disc degeneration is a result of age-related changes. However, this interpretation is based on little more than the circumstantial evidence that the biochemical and morphological features of degeneration increase with age. What remains unexplained is why they occur. Theories of impaired nutrition have not been substantiated. The default explanation is that degeneration is due to programmed cell senescence.

An alternative view is that disc degeneration is caused by mechanical factors. Increasing evidence implicates injury to the vertebral endplate as central to the process. Such injuries can impair the nutrition of the disc or can directly precipitate matrix degeneration. The biomechanical changes of disc degeneration are as much a sign of a response by a connective tissue to injury as due to idopathic age changes. Evidence in favour of mechanical factors is strongly based on laboratory data from cadaver experiments. It should be borne in mind that freezing animal or human cadaver specimens permanently alters disc behaviour compared with normal circumstances.

In spite of this, reasonable evidence can be obtained to support the mechanical concept of disc degeneration, particularly that originating in torsional and compressive injuries. Compressive loading or torsion may produce fracture of the endplate or tear of the annulus respectively, which, in turn, will drive the biological events.

Perhaps the most pivotal development in biomechanical research into disorders of the lumbar spine is the recognition of fatigue failure. Endplate fractures and disc degeneration do not require a single memorable traumatic event: they can occur silently and progressively as a result of repeated, subliminal insults to the disc. This renders mechanical factors difficult to identify in epidemiological studies. The failure to incriminate environmental factors in the aetiology of disc degeneration may lie in the lack of tools with which to detect them. The development of high-resolution imaging techniques may enable detection of tiny fractures of the endplate in vivo.

Based on the available evidence, conventional wisdom dictates that degeneration of the intervertebral disc can be defined as an age-dependent, cell-mediated molecular degradation process under genetic influence that is accelerated primarily by nutritional and mechanical factors, and secondly by toxic or metabolic influences. These factors mediate degeneration by triggering chemical reactions. These changes can affect the morphology of the disc, manifested as evidenced by thickening of the vertebral endplate, cracks and fissures in the matrix, delamination and tears in the annulus, and in the biomechanical function of the disc (Fig. 1). The end result of disc degeneration is characterised by collapse of the intervertebral space and osteophyte formation.

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

References

The proteoglycans of the human intervertebral disc: changes in their biosynthesis and structure with age, topography, and pathology. Spine 1985;10:2271-84.


The large proteoglycans of the human intervertebral disc: changes in their biosynthesis and structure with age, topography, and pathology. Spine 1985;10:2271-84.


