Models for the pathogenesis of stress fractures in athletes

K L Bennell, S A Malcolm, J D Wark, P D Brukner

Although mechanical loading associated with exercise participation can lead to positive bone adaptation, structural damage may result if loading is excessive. Stress fractures represent one form of breakdown in the skeleton. They can be defined as partial or complete fractures of bone that result from the repeated application of a stress less than that required to fracture bone in a single loading situation. While it is apparent that stress fracture development involves repetitive loading and the processes of both bone fatigue, with microdamage accumulation, and remodelling, the exact pathogenesis is unknown. Based on a review of published reports, we shall present two possible models for the pathogenesis of stress fractures in athletes. Understanding the pathogenesis may aid in the identification of factors which place an athlete at increased risk for this sports injury.

Historical perspective
Stress fractures were first described in 1855 by Briethaupt, a Prussian military physician. Observing foot pain and swelling in young recruits unaccustomed to training, he considered the condition to be an inflammatory reaction in the tendon sheaths due to trauma. It was not until the advent of radiographs, 42 years later, that the signs and symptoms were attributed to fractures in the metatarsals. Later reports described similar occurrences in other bones such as the calcaneus, fibula, tibia, femur, and pelvis. Although stress fractures have now been described in almost every bone in the body, they are more common in the bones of the lower extremity.

Most of the published reports on stress fractures derive from military experience. Stress fractures were first noted in civilians in 1921 by Deutschlander who reported six cases in women. They were later described in ballet dancers, weekend outdoor enthusiasts, children, farmers, middle aged homemakers, factory workers, nurses, and even pregnant women. However, they were not recognised in athletes until more than a century following their identification in military recruits.

Pathogenesis of stress fractures
Over the years many theories have been proposed to explain the aetiology of stress fractures. These included spasticity and spasm of the interosseous muscles, impaired circulation, and inflammatory causes such as non-suppurative osteomyelitis. In 1937, Deitelsch and in 1942, Hartley both postulated the theory that stress fractures are related to bone exhaustion, being analogous to fatigue fractures of common metals. This view, with minor modification, is currently accepted today.

The confusion in aetiology has led to a variety of terms used to describe stress fractures. The earliest terms included "march fractures" or "Deutschlander’s fractures" because of the clear association between marching and the onset of symptoms. Some of the other terms used for varying periods of time have been fatigue fracture, crack fracture, spontaneous fracture, insufficiency fracture, pseudofracture, and exhaustion fracture. Stress fracture is the term most popular in current medical usage.

IN VITRO STUDIES OF BONE FATIGUE DURING REPETITIVE MECHANICAL LOADING
Cyclic fluctuating loads may cause bone to undergo gradual mechanical failure, a process known as fatigue. An indication of bone strength during repetitive loading can be gained in vitro by measuring the number of cycles to failure, a variable known as fatigue life. Since the earliest study in 1957 there have been numerous conflicting reports on the fatigue life of human bone. Some found an extremely long fatigue life and others the contrary. Methodological differences make comparisons difficult and probably explain the discrepancy in results. In addition, these in vitro studies treat bone as a purely material substance and cannot take into account factors such as remodelling or muscle activity which can markedly influence bone fatigue.

Nevertheless, after a number of loading cycles, a loss of stiffness, a marked increase in nonlinearity and hysteresis of the stress-strain curve, and a reduction in ultimate strength have been consistently demonstrated in bone specimens subjected to both physiological and hyperphysiological loading regimes. These changes are consistent with the development of microscopic cracks within bone. Under continued loading, these microcracks may propagate and coalesce into "macro-
cracks", resulting in a stress fracture. Evidence to support the existence and progression of microdamage in bone during repetitive loading will follow.

PRODUCTION OF BONE MICRODAMAGE DURING REPETITIVE LOADING

The first reports of microdamage in bone tissue were of cracks in human rib sections. Since then, microscopy and indirect techniques such as acoustic emission signals have revealed the existence of bone microdamage following repetitive loading in vivo. Although many studies employed hyperphysiological strain, microdamage has also been observed in bone subjected to physiological strain. This suggests that microdamage may be a normal phenomenon in humans.

Results of in vivo studies differ, and this may partly be due to the wide variety of loading regimens and measurement techniques employed by investigators and the use of either animal or human models. Some studies failed to show microdamage following repetitive loading, perhaps because the exercise period was insufficient to initiate microdamage or because imaging may not have been sufficiently sensitive. In a group of 98 athletes, high frequency ultrasound was used to assess changes in mechanical properties of human bone indirectly following a 21 km marathon or approximately 30,000 loading cycles. Although the authors concluded that fatigue microdamage did not accumulate, it is difficult to conclude on the basis on ultrasound alone, measured at one point on the tibia, that microdamage was not initiated in this bone.

Conversely, several animal studies have provided substantial evidence of bone microdamage following repetitive loading in vivo. Forced jumping produced cracks apparent on histology of rabbit tibia after 10 days. With continued jumping, these cracks accumulated into fractures in some animals. These findings were confirmed in another rabbit model, as histological microdamage was evident at sites subjected to repeated loading and showing scintigraphic and radiological evidence of tibial stress fracture. Therefore physiological strains may initiate microdamage, corroborating the findings of in vitro studies.

It is possible that a threshold level exists for the onset of microdamage. Significant microdamage was produced in dog forelimbs subjected to three point bending at 1500 or 2500 microstrains for 10,000 cycles, but not when subjected to lower strain or fewer cycles. Frost supports the concept of a microdamage threshold for bone. Based on fatigue, clinical, and pathological studies, he suggests that this threshold is approximately 2000 microstrains, which represents the upper range of physiological values. He also hypothesises that the relation between strain and microdamage becomes exponential at deformations greater than 2000 microstrains. Perhaps in studies where microdamage was not evident, the threshold level for that bone region may not have been reached. Therefore the intense training of many athletes could normally be expected to result in microdamage at maximally stressed bone sites.

FACTORS INFLUENCING MICRODAMAGE PROGRESSION

Several factors have been shown to influence microdamage progression and therefore stress fracture development. Those that increase the likelihood of microdamage include high magnitude stress, greater physiological strain rates, increased number of loading cycles, compressive strain modes, increased loading frequency, and increased Haversian remodelling. The existence of bone microdamage has also been observed in bone subjected to physiological strain. This suggests that microdamage may be a normal phenomenon in humans.

Mechanisms for the prevention of microdamage accumulation

Since the threshold for bone microdamage appears to be within the upper range of normal activities, mechanisms must exist to prevent progression of microdamage. One mechanism is that crack propagation is limited by the osteonal structure of Haversian bone. Failure at weak interfaces by delamination at the cement lines deflects and traps the crack, increasing fracture toughness and extending the fatigue life. Another mechanism is repair of bone damage through the remodelling process. Remodelling is a cyclic process whereby resorption always precedes repair of microscopic cavities in bone. It is performed by teams of osteoclasts and osteoblasts organised into special bone remodelling units. Bone remodelling could repair microdamage in two ways: by directed repair, whereby the remodelling units are directed to the location of damage, or by simple random remodelling of the cortex at a rate designed to keep up with damage accumulation. Of these, directed repair is favoured by most investigators. However, for directed repair to occur, there must be some initiating stimulus. Although the nature of this stimulus is unknown, possibilities include a cellular membrane response resulting from disruption of osteocytes, and the canalicul network, or an electrical response in the Haversian canal cells due to osteon debonding and cracking.

EVIDENCE FOR REPAIR OF BONE MICRODAMAGE BY REMODELLING

Histological repair of microdamage was first documented in 1967. Canine ulnae were overloaded by resection of a portion of the radius. Increased osteoclastic and osteoblastic activity was found in the fatigued damaged regions of the ulnar cortex. These results were corroborated in a later study using a similar canine model. Another in vivo study involved repetitive loading of the radius and ulna of dogs using physiological strain regimes. Light and scanning electron microscopy showed that there were 44 times as many microcracks in direct association with resorption spaces than expected by chance alone. However, while these studies show an association between
microdamage and bone remodelling, they do not prove a cause and effect relation.

Whether microdamage precedes or follows bone remodelling is unclear, as studies which investigate the temporal relations between microdamage production and bone remodelling have provided conflicting results. Mori and Burr reported findings consistent with the hypothesis that microdamage is followed by bone remodelling. By loading both limbs identically but for different periods, these investigators were able to show a significant increase in new remodelling events following the generation of microdamage, which occurred preferentially in fatigue damaged regions. However, there were still three times as many resorption spaces as microcracks, suggesting that factors other than microdamage also initiate remodelling. These factors may include systemic and reproductive hormones, dietary factors, and bone strain arising from mechanical loading.

Conversely, some human studies suggest that microdamage occurs at pre-existing sites of accelerated remodelling, where osteoclastic resorption weakens an area of bone and subjects it to higher strains, before the addition of new bone by osteoblasts. In a temporal series of stress fracture biopsies mainly from the upper tibial cortex in humans, initial histology revealed accelerated cortical resorption. Although no microfracture was seen at this stage, a thin crack was evident in many of the specimens a week later, followed by osteoclastic activity and new bone formation. However, these studies do not reveal the exact time course of changes in response to loading. It is possible that microdamage was present before osteoclastic resorption but was histologically undetectable.

The study by Li et al employed an exercising rabbit model to assess sequential pathological changes in the internal structure of the tibia during a 10 week period. Within the first week, osteoclastic resorption cavities appeared in the tibial cortex and interstitial lamellae, followed, in the second week, by small cracks at the cement line of the Haversian system, together with obvious osteoclastic resorption. By the third week, incomplete fracture of the tibial cortex was found in some specimens. Over the remaining six weeks, the resorption cavities gradually filled with bone and converted to Haversian bone. One specimen developed a cortical fracture. Thus most tibiae adapt successfully to changes in bone strain from repetitive loading through internal remodelling, but fractures may appear if excessive stress continues in a tibia weakened by osteoclastic resorption.

Indirect support for the hypothesis that accelerated remodelling, either local or generalised, may be pathogenic for stress fracture is provided by a prospective study of bone turnover in 104 male military recruits. Plasma hydroxypyroline (a non-specific indicator of bone resorption), measured in the first week of a training programme, was significantly higher in those who subsequently sustained stress fractures than in those who remained uninjured. This supports the concept that increased bone turnover may be a stimulus for stress fracture development.

**ROLE OF MUSCLES IN STRESS FRACTURE DEVELOPMENT**

The role of muscle contraction in stress fracture development is unclear and may vary depending on the site of fracture. Nevertheless, the contribution of muscle to stress fracture pathogenesis is merely to alter the mechanical environment of bone. Some investigators consider that muscles act dynamically to cause stress fractures by increasing bone strain specifically at sites of muscle attachment. For example, contraction of the calf muscles may contribute to calcaneal stress fracture through the Achilles attachment. Greater muscle mass with greater ability to generate force would thus be associated with an increased risk for stress fracture. However, stress fractures often occur at sites unrelated to muscle attachment. Since muscles also act to attenuate and dissipate forces applied to bone, muscle fatigue or muscle weakness could predispose to stress fracture by causing an increase and redistribution of stress to bone. In the military, leg power was not associated with stress fracture occurrence, although the testing method was relatively crude and non-specific. However, in both recruits and female athletes, a larger calf muscle circumference has been found to be associated with significantly fewer stress fractures. This supports the hypothesis that muscles act to protect against rather than cause stress fractures.

There have been no studies comparing muscle mass or muscle strength, particularly peak force production and fatigueability, in athletes with and without stress fractures. However, Grimston et al found that during the latter stages of a 45 minute run, women with a past history of stress fracture recorded increased ground reaction forces, whereas these forces did not vary during the run in the control group. These investigators surmised that this may indicate differences in fatigue adaptation and muscle activity which could contribute to stress fracture development.

**Models for the pathogenesis of stress fractures**

It is apparent that stress fracture development is a process and not an event, and represents the end product of bone fatigue as a result of repetitive loading. The processes of microdamage accumulation and bone remodelling play an important part in stress fracture pathogenesis. Based on the literature to date, two possible scenarios, not necessarily mutually exclusive, may explain the development of a stress fracture. These are depicted diagrammatically in the figure.

**PRIMARY MICRODAMAGE HYPOTHESIS**

Bone strain from repetitive loading initiates the production of microdamage at particular sites which are maximally stressed. A remodelling response is stimulated at the damaged site in order to affect repair. In physiological situations, a balance exists between these two proc-
Stress fracture pathogenesis

Diagrammatic representation of possible mechanisms for stress fracture development: primary microdamage (MDx) hypothesis and primary remodelling hypothesis.

Stresses and the microdamage is adequately repaired. The development of a stress fracture is thought to occur when microdamage production exceeds repair. This may occur because the microdamage is too extensive to be repaired by normal remodelling, because local remodelling, particularly bone formation, is impaired, or because of a combination of these factors.

**Primary Remodelling Hypothesis**

Accelerated bone remodelling, which may be a local or a generalised skeletal phenomenon, is the initiating stimulus to stress fracture development in this hypothesis. Various factors may result in accelerated bone remodelling including genetics, bone strain arising from exercise, systemic or reproductive hormones, and dietary intake. Since osteclastic resorption always precedes formation in the remodelling process, there is a lag time in which the bone is in a weakened state. Microdamage may occur at these focal areas of weakness. As with the primary microdamage hypothesis, a stress fracture may develop if loading continues. The difference between these two hypotheses lies in whether the process of remodelling precedes or follows microdamage production.

**Clinical Relevance**

Numerous factors have been suggested as increasing the risk of stress fracture. These include training "errors", low bone density, inadequate dietary intake, menstrual disturbances, and biomechanical variants. From an epidemiological perspective, risk factors are markers that can be used to identify individuals who are most likely to sustain a stress fracture. However, risk factors are not the cause of this injury, but are factors that might directly or indirectly affect the mechanical environment of bone or the remodelling process, thereby increasing the rate of microdamage formation or decreasing the repair of this damage, or both. For example, higher weekly running mileage will increase the number of bone loading cycles, low bone density will decrease the ability of bone to withstand the applied load, and inadequate dietary calcium may accelerate remodelling or impede bone formation during repair of microdamage. Thus an understanding of stress fracture pathogenesis is necessary to identify possible risk factors and to establish preventative measures.

**Summary**

It would seem that the development of a stress fracture results from unsuccessful adaptation of bone to a change in its mechanical environment caused by repetitive loading. It involves the physiological processes of microdamage production and remodelling. Whether the initiating factor is microdamage production or activation of remodelling through direct effects of strain is unclear. The remodelling process involves both the removal of bone which has become fatigue damaged or is extraneous to the requirements of the new loading environment, and the addition of new bone in a manner that is best suited to withstand the new mechanical strain. Normally this process is well modulated and does not cause symptoms. If the amount of bone removed is not sufficient to unduly weaken bone structure and the addition of new bone occurs sufficiently rapidly to correct any weakness before failure occurs or to repair microdamage, the process will successfully lead to a bone with appropriate material strength and geometry to withstand the new strain environment. However, if there is imbalance between bone removal and replacement, together with accumulation of microdamage, signs and symptoms of a stress fracture may result. Any factors which influence bone load, bone strength, or remodelling may have the potential to result in a stress fracture. Attention should be paid to the identification of these factors in an attempt to prevent this overuse injury in athletes.


22 King AI, Evans FG. Analysis of fatigue strength of human cortical bone by the Weibull method. Digest 7th International Conference of Medical and Biological Engineering, Stockholm: 1967.


