Classification des fractures des os longs : concepts et utilité

Rod Fleury, Thierry

This publication URL: https://archive-ouverte.unige.ch/unige:89579
Publication DOI: 10.13097/archive-ouverte/unige:89579

UNIVERSITÉ
DE GENÈVE
FACULTÉ DE MÉDECINE

DOCTORAT EN MEDECINE

Thèse de :

Thierry ROD FLEURY

originaire de Jorat-Menthue (VD), Suisse

Intitulée :

**Classification des fractures des os longs -
Concepts et utilité**

La Faculté de médecine, sur le préavis du Comité directeur des thèses, autorise l'impression de la présente thèse, sans prétendre par là émettre d'opinion sur les propositions qui y sont énoncées.

Genève, le 25 octobre 2016

Thèse n° **10823**

Henri Bounameaux
Doyen

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives à la présentation des thèses de doctorat à l'Université de Genève".

2ème version

UNIVERSITÉ DE GENÈVE

UNIVERSITÉ DE GENÈVE
FACULTÉ DE MÉDECINE

Section de *médecine Clinique, Fondamentale, ou Dentaire*
Département : Chirurgie
Service : Orthopédie et traumatologie de l'appareil moteur

Thèse préparée sous la direction du Professeur Pierre J. Hoffmeyer

" CLASSIFICATION DES FRACTURES DES OS LONGS - CONCEPTS ET UTILITE "

Thèse
présentée à la Faculté de Médecine
de l'Université de Genève
pour obtenir le grade de Docteur en médecine
par

Thierry ROD FLEURY

de
Jorat-Menthue, VD

Thèse n° 10823

Genève

2015

UNIVERSITÉ
DE GENÈVE

Formulaire à joindre aux exemplaires de la thèse à remettre pour le dépôt légal

THESE

Informations indispensables à dactylographier

Nom et Prénom : ROD FLEURY Thierry

Adresse : Ruelle Jean-Michel BILLON 4, 1205 Genève

Faculté : Médecine   Département (selon structures officielles) : Chirurgie

Directeur de thèse : Professeur Pierre J. Hoffmeyer

Références bibliographiques :

Classification des fractures des os longs - Concepts et utilité

European Surgical Orthopaedics and Traumatology
The EFORT Textbook
G. Bentley (ed.), Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 2014(1):115-37

Résumé :
Le traitement des fractures des os longs fait partie intégrante des tâches quotidiennes d'un service d'orthopédie et peut représenter une part importante de son activité. Le choix de la stratégie thérapeutique est influencé par de nombreux facteurs plus ou moins objectifs allant de l'expérience de l'opérateur aux recommandations basées sur des preuves scientifiques, en incluant des facteurs liés au patient et aux ressources du service. Cependant, la décision part toujours initialement de la fracture, qui est décrite, puis catégorisée et incluse dans une classification.
L'objectif de ce travail est de développer le concept de classification des fractures en soi, d'en identifier les caractéristiques, l'utilité et les limitations, et d'évoquer les pistes de développements futurs qui devraient permettre d'améliorer la prise en charge des patients victimes de fracture.

Signature du doctorant :   Visa du directeur de thèse :

Thèse de Doctorat en Médecine

Thierry Rod Fleury
Sous la supervision du Professeur Pierre J. Hoffmeyer
Service d'orthopédie et traumatologie de l'appareil locomoteur, Hôpitaux Universitaires de Genève

Classification des fractures des os longs

Concepts et utilité

Table des matières

Introduction et Résumé

Le traitement des fractures des os longs fait partie intégrante des tâches quotidiennes d'un service d'orthopédie et peut représenter une part importante de son activité. Le choix de la stratégie thérapeutique est influencé par de nombreux facteurs plus ou moins objectifs allant de l'expérience de l'opérateur aux recommandations basées sur des preuves scientifiques, en incluant des facteurs liés au patient et aux ressources du service. Cependant, la décision part toujours initialement de la fracture, qui est décrite, puis catégorisée et incluse dans une classification. Celle-ci sert de dénominateur commun aux lésions de même type, impliquant ainsi un traitement identique afin d'obtenir *a priori* un résultat optimal pour le patient.

L'objectif de ce travail est de développer le concept de classification des fractures en soi, d'en identifier les caractéristiques, l'utilité et les limitations, et d'évoquer les pistes de développements futurs qui devraient permettre d'améliorer la prise en charge des patients victimes de fracture.

Il semble naturel à l'esprit humain d'organiser les choses qu'il a nommées en groupes d'objets partageant des caractéristiques communes. Une telle organisation permet de simplifier le monde nous entourant dans le but de mieux le comprendre, de communiquer avec les autres et de guider nos actions.

Tout aussi naturellement, les médecins décrivent et regroupent depuis toujours les fractures selon diverses systèmes de classification dont la structure hiérarchique inclut des types et sous-types. Chaque sous-type comporte l'intégralité des caractéristiques du type initial, plus quelques caractéristiques qui lui sont propres. Ces systèmes de classification de fractures visent quatre objectifs afin de se rendre idéaux pour le médecin.

Premièrement, donner un nom descriptif à la fracture, afin de la différencier d'une autre et de savoir de quoi l'on parle.

Deuxièmement, regrouper et ordonner les fractures ayant des caractéristiques communes. Les groupes ainsi définis peuvent être purement descriptifs, ou plus généralement comporter un ordre supposé ou avéré de sévérité ou de complexité. Ainsi, la communication entre professionnels est simplifiée, et une base de travail pour la recherche est mise en place.

Troisièmement, prédire les résultats. Une classification atteignant cet objectif devrait permettre d'entrevoir *a priori* les résultats de l'évolution naturelle d'une fracture ou des traitements entrepris. Ceci serait d'un grand bénéfice pour le patient et son chirurgien, qui pourraient choisir le traitement optimal dès le diagnostic de la fracture. Cependant, aucun système de classification de fractures ne permet à l'heure actuelle de prédire avec fiabilité les résultats.

Quatrièmement, corolaire du troisième objectif, guider les actions. Pour autant qu'une classification puisse prédire les résultats de façon fiable, et qu'elle ait certaines qualités (validité, fiabilité, reproductibilité) qui seront décrites ci-après, chaque position en son sein devrait impliquer le choix d'un traitement optimal et ainsi guider la prise de décision. Malheureusement, compte tenu de l'impossibilité de prédire les résultats et de l'influence d'une multitude d'autres facteurs sur l'évolution des fractures, aucune classification actuellement disponible n'atteint cet objectif.

Malgré leurs limitations, les classifications de fractures ont une utilité évidente tant dans le domaine de la recherche que dans la pratique au quotidien, puisqu'il est possible de se faire une image mentale de la plupart des fractures sur la simple base de leur nom au cours d'une conversation, même sans radiographie à disposition.

Quatre grands types de classifications peuvent être identifiés : les systèmes spécifiques aux fractures, les systèmes spécifiques aux patients, les systèmes génériques (dits « universels »), et les classifications concernant les atteintes des tissus mous associées aux fractures.

Les systèmes spécifiques aux fractures décrivent les fractures touchant une zone anatomique bien particulière et exclusive. La classification de Schatzker, par exemple, décrit les fractures du plateau tibial, et ne peut s'appliquer qu'à cette localisation anatomique précise.

Les systèmes spécifiques aux patients s'appliquent à une certaine catégorie de patients, tels que les enfants ou les patients atteints de cancers, mais sans restriction de localisation anatomique. Un bon exemple est la classification de Salter-Harris qui décrit les fractures pédiatriques touchant les plaques de croissance, quel que soit l'os concerné.

Le seul système universel actuellement disponible est celui de l'AO (Arbeitsgemeinschaft für Osteosynthesefragen). L'objectif d'un tel système est de classer les fractures de n'importe quel os en appliquant toujours la même méthodologie logique. Il s'agit d'un système de codage alphanumérique dont chaque élément donne une information incluant l'os considéré, le segment touché, et le type de fracture avec éventuellement des précisions sur des sous-types. L'avantage est de permettre de classer toutes les fractures possibles avec un seul système, en évitant l'écueil des éponymes, toutefois au prix d'une certaine complexité et d'un manque d'aspect instinctif.

Enfin, les classifications des lésions des tissus mous associées aux fractures permettent de décrire le traumatisme subit par les tissus avoisinant l'os, sans décrire la fracture elle-même. Il s'agit d'informations complémentaires importantes pour la prise en charge et le pronostic, qui viennent s'ajouter à une autre classification décrivant la fracture en elle-même. Les plus connues sont la classification de Gustilo et Anderson des fractures ouvertes, et celle de Tscherne des fractures fermées.

En tant qu'outils de travail, un système de classification doit avoir sept qualités afin qu'il soit idéal, et non pas rendu inutile par des défauts qui lui sont inhérents.

La validité est la capacité du système à décrire précisément le véritable état de la chose observée. Il s'agit ici de la corrélation entre la classification et la réalité de la fracture. Celle-ci n'étant pas accessible en raison des biais de l'observation, il s'agit donc de comparer l'outil en question avec un *gold-standard*. Cet étalon n'est toutefois pas toujours facilement disponible.

La reproductibilité est définie par le fait qu'une fracture donnée soit classée toujours de la même façon par différents observateurs. On parle aussi de fiabilité inter-observateur.

La répétabilité implique quant à elle qu'un même observateur classe la même fracture toujours de la même façon à plusieurs occasions différentes. On parle également de fiabilité intra-observateur.

La tout-inclusivité et l'exclusivité mutuelle impliquent que chaque fracture possible peut être classée dans une et une seule catégorie de la classification relative. Un système est imparfait si une fracture ne peut pas y être classée, ou si elle peut se trouver dans deux catégories en même temps.

Finalement, la logique et l'utilité clinique se définissent d'elles-mêmes. Un système qui ne serait pas logique ne serait que source de confusion, et l'absence d'utilité clinique le rendrait simplement inutile.

En raison de la difficulté de mesurer sa validité, un système de classification devrait avoir au moins de hauts degrés de fiabilité intra- et inter-observateurs. La méthode la plus simple pour mesurer ces paramètres est de déterminer l'accord brut, c'est-à-dire le pourcentage de fois où les observateurs sont tombés d'accord dans leurs évaluations. S'ils sont concordants sur 75 cas au cours de 100 évaluations, l'accord brut est donc de 75%. La limite de cette méthode est qu'elle ne prend pas en compte le facteur chance, le fait que les observateurs aient pu être d'accord par chance. Ainsi, un nouvel outil a été introduit, le facteur statistique Kappa (κ), qui mesure l'accord au-delà du facteur chance.

K est désormais le coefficient statistique le plus utilisé afin d'évaluer la fiabilité des systèmes de classification de fractures actuels. Il présente toutefois des limitations, notamment dans son interprétation. Son intervalle de valeur va de -1 à 1, avec -1 correspondant à un désaccord total, 0 à un accord dû uniquement à la chance, et +1 à un accord total. Il n'y a cependant pas de valeurs limites mathématiquement définies entre ces trois bornes afin d'évaluer la qualité de l'accord, même s'il semble logique que plus haut est le κ, meilleur est l'accord. Différents auteurs ont donc choisi arbitrairement des valeurs limites de qualité de κ, sans qu'il y ait de consensus global.

De nombreux chercheurs ont donc évalué la fonctionnalité des différentes classifications de fractures, tout d'abord en termes de fiabilité intra- et inter-observateurs. L'étape suivante aurait été d'évaluer leur utilité clinique, ce qui n'a malheureusement pas pu être fait puisque, hormis quelques rares exceptions, la plupart de ces classifications obtiennent des scores de fiabilité tout à fait décevants. Où se situe le problème, pourquoi y a-t-il si peu de classifications fiables ?

Tout d'abord, certaines classifications comportent des défauts inhérents de conception, telle que la classification des fractures du plateau tibial de Hohl et Moore, qui n'est pas tout-inclusive. Certaines fractures ne pouvant y être classées, les examinateurs choisissent la catégorie la moins fausse plutôt que la meilleure, ce qui crée de l'imprécision et de la variabilité. Un autre problème de conception survient lorsque la classification est basée sur des paramètres non radiologiques tels que la qualité osseuse, ou la stabilité de la fracture ou des implants. Ces paramètres sont très délicats à évaluer, notamment en raison de l'absence de critères radiologiques objectifs le permettant. De l'incertitude et de la variabilité en découlent ainsi.

La complexité d'un système de classification est un autre facteur que l'on pourrait raisonnablement considérer comme source de problème. Un système plus complexe devrait être plus difficile à comprendre et donc amener l'observateur à hésiter dans ses choix lors du processus de classification. Cela n'est toutefois pas vérifié, puisque certaines classifications complexes ont une meilleure fiabilité que d'autres plus simples, et que les essais de simplification de classifications complexes n'ont pas procuré d'amélioration de fiabilité.

De même, il pourrait apparaître logique que l'expérience de l'observateur joue également un rôle dans la fiabilité de la classification de fractures. Cependant, de nombreuses études ont montré que cela n'est pas le cas, et que ce facteur n'a pas d'influence sur la fiabilité d'une classification.

Finalement, s'il semble évident que la qualité des examens radiographiques devrait influencer la capacité d'un observateur à classer correctement une fracture, notamment si les lignes de fracture sont difficilement identifiables, plusieurs études ont montré que l'amélioration de la qualité des radiographies standards, voire même l'identification préalable des traits de fracture sur celles-ci, ne permet pas d'atteindre un κ satisfaisant. Contrairement au CT 2D qui ne permet pas non plus d'améliorer la fiabilité des classifications, l'utilisation de technologies plus avancées telles que la reconstruction 3D de CT et la stéréo-visualisation de CT 3D permet une amélioration significative de la fiabilité de classification des fractures ainsi traitées. Il reste néanmoins à déterminer si les coûts et le temps accrus liés à ces technologies ont un intérêt clinique pour les patients, et économique pour les prestataires de soins.

Les classifications de fractures actuelles sont confrontées à de nombreuses limitations qui les empêchent de remplir convenablement deux de leurs principaux rôles : guider les actions et en prédire les résultats. Ceci jette ainsi un doute sur de nombreuses études recommandant un certain traitement ou implant en se basant sur la classification de la fracture concernée. Malgré tout, nous continuons d'utiliser ces classifications dans notre pratique quotidienne car il n'y a tout simplement pas de meilleur outil à disposition pour communiquer et prendre des décisions efficacement et rapidement.

Il est désormais nécessaire d'améliorer le concept de classification des fractures dans sa globalité, sans rester focalisé sur la radiographie osseuse uniquement. De nombreuses autres variables, par exemple

liées aux tissus mous ou aux facteurs socio-économiques du patient, doivent certainement être prises en compte et incluses dans les classifications. Des groupes d'experts devront définir formellement certains concepts encore imprécis tels que la stabilité, le déplacement ou la comminution des fractures. Ils devront également s'entendre sur la méthodologie des études qui évalueront la fiabilité des systèmes de classification, ainsi que sur des valeurs limites définissant une fiabilité acceptable ou non.
Les développements techniques en imagerie et en technologies de l'information devraient permettre d'améliorer tant notre compréhension des fractures que la fiabilité de classification de celles-ci. Le développement d'outils informatiques facilitant ou même automatisant la classification est imaginable à partir du moment où les éléments caractérisant les différentes fractures auront été précisément définis.
En conclusion, le monde de la classification des fractures subit actuellement une indispensable évolution. Que ce soit par la création de nouveaux systèmes ou par l'amélioration de ceux existant, l'avènement de classifications fiables est nécessaire à l'amélioration de la qualité des soins offerts aux patients, mais également à la détermination précise des coûts des traitements.

La version anglaise de ce travail a été publiée comme chapitre du livre de référence :

European Surgical Orthopaedics and Traumatology
The EFORT Textbook
G. Bentley (ed.), Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 2014(1):115-37

Cette thèse est dédiée à Mapi, Elina, Eulalie et Eole, pour leur amour, leur patience et leur soutien indéfectible.

Classification of long bone fractures

Thierry Rod Fleury and Richard Stern
Division of Orthopaedics and Trauma Surgery
University Hospitals of Geneva

Introduction

Since human beings acquired speech, they named things. Emmanuel Kant thought that the adult human mind naturally organizes its knowledge of the world in groups of objects sharing the same name. The purpose of such organization is to simplify the surrounding world in order to better understand it, communicate with others, and guide actions.

Taxonomy is the science of naming and classifying items, originally concerning only organisms but later extended to the classification of any concept or thing that can be classified. The basic unit is named *taxon*, and these units are arranged in a hierarchical structure, usually with a type-subtype relationship. The subtype has all the properties of the parent type, plus one or more additional properties characteristic of itself. For example, the living world could be divided into three kingdoms: animals, plants and bacteria. A mammal is an animal but not all animals are mammals, and a cow is a mammal but not all mammals are cows.

Fracture classification systems based on the same reasoning have probably existed for nearly as long as people have identified fractures. This work reviews the concept of fracture classification, with the history, purposes, types, strengths and limitations of the current classification systems.

Classification systems: What are they used for?

In all its fields of application, scientific or otherwise, the first purpose of taxonomy is to describe and name things. A name permits the differentiation of one object from another, and to identify it in order to work with it. In a fracture classification system, the beginning is also the setting of names in order to understand what we are talking about.

The second purpose is to group and order the objects. Groups are defined according to their description and to a choice of common characteristics of the objects composing them. The choice of the common descriptors of the group can be completely empirical, or it can obey some logical or scientific criteria. The interesting point is the variable nature of these groups which can be modified according to scientific progress or expert opinion. For example, the classification of nature (*Systema Naturae)* by Carolus Linnaeus in 1735 (1) included three *kingdoms* (mineral, vegetable and animal) with organisms classified by shared physical characteristics. One of the most recent systems was invented in 1990 by C. Woese (2) and includes three domains (bacteria, archaea and eukaria) according to the Darwinian principle of evolution. Fungi were part of the kingdom of plants in Linnaeus' system, but they are now a kingdom in itself, included in the domain of eukaria.

The defined groups are then ordered in a hierarchical manner, usually with increasing complexity of characteristics as the progression goes down along the branches of the tree-diagram drawn by this classification, the typical example being the phylogenetic Tree of Life. This is a branching diagram or "tree" showing the inferred evolutionary relationships among various biological species or other entities based upon similarities and differences in their physical and/or genetic characteristics. Fracture classification systems are often based on the same principle. The choice of the characteristics of each group can also be empirically based upon physical properties like fracture lines or fracture patterns, or it can fit with scientific data like outcome prognosis or biomechanics. Examples are the Schatzker classification of proximal intraarticular tibia fractures which is based upon fracture lines and patterns, or the Danis-Weber classification of malleolar fractures which was developed according to the fracture mechanism.

A third purpose of fracture classification systems is to predict outcomes. This was one of the principles that led to the development of the "Comprehensive Classification of Fractures" by Professor

Maurice Müller in 1990 (3). Unfortunately, at present, no fracture classification system can reliably assist in predicting outcomes following the most common fractures. This is explained by many factors: The natural outcome of a fracture must be known, the impact of different interventions on the natural outcome must be studied, and above all the classification has to prove itself to be valid, reliable, and reproducible. Theoretically a classification system which had all these qualities would be of tremendous benefit to patient and surgeon for it would allow for an expectation of the outcome at the time of fracture, and thus help decide upon the most effective treatment.

That leads to the fourth purpose of classification systems, guiding actions. This feature is not common to all classification systems, but only to those implying an action in reaction to a *diagnosis*. Descriptive-only systems are devoid of this feature as there is no consequence to a description. For example, the Thorne system of plant classification is a purely descriptive classification system, and no specific action is suggested in reaction to the position of a plant in the system. In contrast, most fracture classification systems were designed by their inventors in order to guide practicing orthopaedic surgeons in the treatment of their patients, based upon the severity, complexity, mechanism of injury, or outcome of the fracture. Such classifications have also to be valid, reliable and reproducible because the absence of these qualities implies an unpredictability of the outcome, and so the systems become useless.

A little bit of history

Fractures have existed for as long as human being have sustained trauma. There is no doubt that shamans and healers in the Prehistoric Ages knew how to recognize fractures and thus attempted to treat them with their limited means. The most ancient text about general and osteologic surgery known today is the Edwin Smith Papyrus (4), which is dated from the beginning of the XVIIIth Egyptian Dynasty (about 1567 B.C.). The author (some think Imhotep himself was the author, although others have claimed they were written and edited by at least three different authors) describes 48 cases of contusions, wounds and fractures ordered by topography from the head and face to spine and long bones. Treatments are advocated for each ailment, according to ancient Egyptian medical principles.

The skillful Egyptian healer knew the natural outcome of the diseases; those that had a good natural outcome were “to be treated”, those that were uncertain had “to be fought”, and those that had a bad natural outcome with or without treatment were “not to be treated”. In the Edwin Smith Papyrus the closed fractures (“ailments to be treated”) are clearly distinguished from open fractures (“ailments not to be treated”), as open fractures were synonymous with early death in ancient Egypt.

In his book “*On fractures*” (5), Hippocrates (*ca*. 460 BC - *ca.* 370 BC) shows that the ancient Greeks also distinguished between closed and open fractures, and the guidelines of treatment depending upon this classification. He also compares elbow with knee dislocations, and gives different methods of reduction for each type of dislocation. In his treaty “*De Medicina*” (6), the Roman encyclopedist Aulus Cornelius Celsus (*ca.* 25 BC - *ca.* 50 AD) demonstrates an astonishingly wide knowledge of long bones fractures. The fractures are classified in degrees of severity based on several characteristics, and guidelines for treatment are provided for each situation. He differentiates between undisplaced and displaced fractures; diaphyseal fractures and fractures of the extremities of the bone (which are more painful and more difficult to cure); simple and transverse fractures which are considered less severe than oblique or comminuted ones; and closed and open fractures which are treated in different ways. Their prognoses are discussed. In the forearm and the leg, he differentiates single bone fractures from two bone fractures. He mentions also limb shortening and soft tissue problems associated with the fracture. Celsus’ writings were in use for many centuries, in fact until the end of the Renaissance.

In the modern era, still before the advent of roentgenography, some authors designed their own fracture classification systems based on the clinical appearance of the affected limb. In the eighteenth century, the Pott’s fracture (7) described a distal tibia and fibula fracture with a varus deformity. In the nineteenth century (1814) Giovanni Batista Monteggia described a fracture of the proximal third of the ulna in association with an anterior dislocation of the radial head (8). At the same time, any fracture of the distal radius with a dorsal deformation in a *dinner fork* shape was classified as a Colles (Colles-

Pouteau) fracture (9), and was treated according to Colles' advice: correct the deformity and immobilize the limb.

With the advent of radiography at the end of the nineteenth century, orthopaedic surgeons had an almost direct view on their subject: the bone. Fracture classification systems multiplied and came into common usage, especially in the medical literature. The changes that took place with a better understanding of fractures brought by radiographs dramatically altered the way of classifying fractures. The patient and clinical status were disregarded, and almost all systems of classification developed from that time were based exclusively on fracture characteristics that were visible and measurable on plain radiographs. Countless fracture classification systems have been described, and most of them have been forgotten. For example, Schepers found 49 systems of calcaneal fracture classification based upon plain radiographs, of which 30 were determined to be of historical significance only (10). But some "old" fracture classification systems are still in common use presently, such as the Garden classification of femoral neck fractures (11) and the Neer classification of proximal humerus fractures (12) being among the most famous.

In the last 30 years the development of new radiologic technologies, principally computed tomography (CT), ushered in a new era in fracture classification systems. Initially, most investigators tried to apply CT data to existing fracture classification systems previously designed for plain radiographs only, in order to improve the performance of these systems. However new classification systems based on CT technology itself have also been designed, perhaps the most famous are the Sanders (13) and the Zwipp (14) classifications of calcaneal fractures. As of the present day there is no publication of a classification system of long-bone fractures based specifically on magnetic resonance imaging (MRI).

Some authors also returned to consideration of the patient as a whole, and to examine non-radiographic factors that could influence the choice of treatment and the outcome of the fracture (15-18). The extent of soft tissue injury, the patient's age and comorbidities, the presence of other traumatic injuries (musculoskeletal or not), and even the social and psychological status of the patient, are some of these factors. Some of them, like soft tissue injury, are the subject of separate classification systems. However none of these factors are part of a radiological fracture classification system.

Recently some authors have reasonably questioned the validity and the usefulness of the fracture classification systems in use at present (19). Currently, research is oriented toward verifying the validity of the classification systems, improving existing systems and developing brand new valid tools.

Types of fracture classification systems

Fracture classification systems can be grouped into four main categories: [1] Fracture-specific systems are designed to describe fractures of a precise and exclusive skeletal location. For example, radial head fracture classification systems apply only to the radial head; [2] Patient-specific systems apply only to a certain category of patients, such as children, or patients with cancer, but are not restricted to a specific bone; [3] The aim of generic or universal classification systems is to classify any fracture of any bone by always applying the same logical methodology. This is basically a numerical coding system; and [4] The fourth group of classification systems deals with the soft tissue injury associated with the fracture, rather than with the fracture itself.

The objective here is not to describe every possible fracture classification system, but rather to discuss the concept of fracture classification. However, some examples of every kind of system will be useful to understand the following discussions.

Fracture-specific classification systems

The Neer classification of proximal humeral fractures (12) is a descriptive classification system described by Charles Neer in 1970. It is still widely used by orthopaedic surgeons around the world, and is one of the most studied fracture classification system. The classification is based on the number of fracture “parts” on plain radiographs **(Figure 1)**. A part is defined as a bone fragment that is displaced more than 1 cm or angulated greater than 45 degrees.

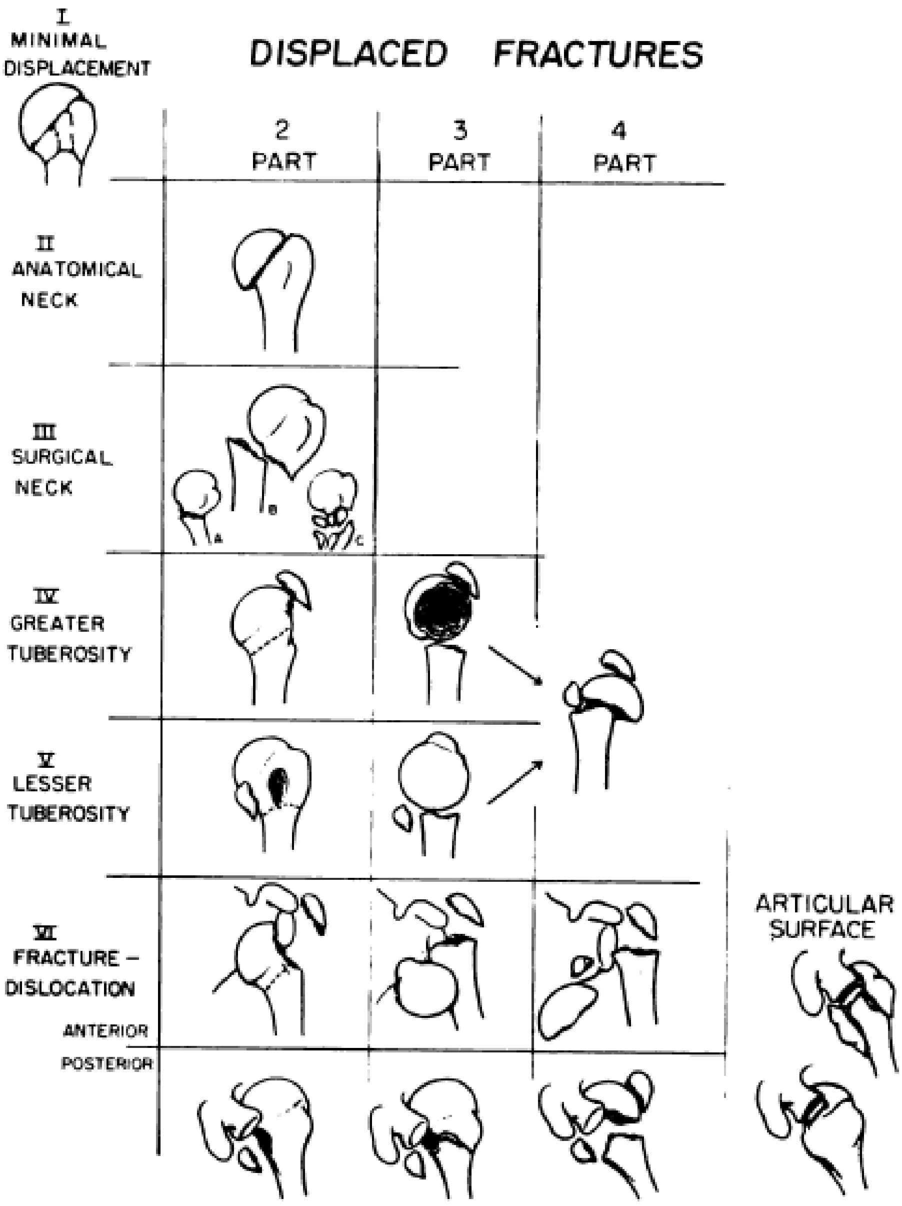

Figure 1 : Neer classification of proximal humerus fractures
Source : Neer CS. Displaced proximal humeral fractures. Part I: classification and evaluation. *J Bone Joint Surg (Am)* 1970; 52:1077-1089.

Neer grouped fractures into five categories: one-part (non-displaced fractures, meaning that no fragment meets the criteria to be a part whatever the number of fracture lines); two-parts (usually the head separated from the shaft, or a greater tuberosity fracture); three-parts; four-parts; and articular surface fractures (usually the head-split type). The challenge of the Neer system is for the observer to correctly and precisely identify every fracture fragment and measure their linear and angular displacements in order to determine which ones are considered as parts.

The Garden classification of femoral neck fractures (11) was originally published in 1961. It is based on the anteroposterior radiographic appearance of femoral neck fractures in varying stages of displacement before reduction. A correlation is made between the radiographic appearance and the anatomic reality of the fracture, which is then grouped into four stages. Stage I are incomplete fractures; Stage II are complete fractures without any displacement; Stage III are complete fractures with partial displacement, the fragments still being attached by the posterior capsule; and Stage IV which are complete fractures with full displacement. These descriptive stages of fractures are ordered in increasing severity of displacement and of expected difficulty in reduction. For each stage of fracture, specific guidelines were proposed on how to achieve a reduction. At that period, internal fixation was the proposed treatment for every fracture. Common use of this classification altered it in some ways. Most of the fractures currently classified as Grade I are complete fractures with valgus impaction, but according to Garden these fractures should be considered as Grade III (complete fracture with partial displacement) since Grade I are only incomplete fractures. It is unclear from the literature when and how this complete valgus impacted fracture came to be classified as a Garden stage I (17). The notion of fracture stability and risk of femoral head osteonecrosis also changed. Garden considered only stage IV fractures as unstable (stability was assured in stage III by the posterior capsular attachment), and linked the risk of complications with the quality of the reduction. In present day usage, Garden stage III and IV are considered unstable with a high risk of avascular necrosis, and thus are commonly treated by arthroplasty in elderly patients.

The Schatzker classification of tibial plateau fractures (20,21) is one of the most widely used descriptive classification system for intraarticular proximal tibia fractures. The fractures are distributed in six types in accordance with the location of the fracture lines and the presence or absence of a depression of the articular surface. The amount of displacement or depression is not taken into account to classify fractures. Type 1 are pure split fractures of the lateral plateau. Type 2 are split fractures of the lateral plateau associated with a depression fracture of the lateral articular surface. Type 3 are pure articular depression fractures of the lateral plateau, without a split fracture. Type 4 are medial plateau fractures, either split or depression, with or without a fracture of the tibial spines. Type 5 are bicondylar fractures, with continuity between the diaphysis and the central metaphysis. Type 6 are tibial plateau fractures with dissociation between the tibial metaphysis and diaphysis.

The Danis-Weber classification (commonly called simply “Weber”) of ankle fractures (22) describes malleolar fractures according to the location of the fibular fracture line, from which the mechanism of injury and the damage to the syndesmosis can be deduced. Weber A fractures are located distal to the syndesmosis, and are typically supination injuries. Weber B fractures are located at the level of the syndesmosis, and are usually pronation injuries. In Weber C fractures, the fracture line is proximal to the syndesmosis; the mechanism of injury is thought to be pronation and external rotation of the foot. The extreme example of a Weber C fracture is the Maisonneuve fracture, in which the fracture line is located right below the head of the fibula.

The Vancouver classification of periprosthetic proximal femoral fractures (23) is peculiar in that it concerns fractures of the proximal femur in patients with an ipsilateral hip arthroplasty. It describes the location of the fracture line in relation to the prosthesis, and takes account of the stability of the femoral component and of the quality of the surrounding bone stock, in order to guide treatment decisions. The femur is divided into three zones: Zone A is the proximal metaphysis, Zone B the diaphysis around the femoral component, and Zone C is the distal diaphysis below the femoral component. Vancouver A fractures involve the greater or lesser trochanters, but do not extend into the diaphysis. Vancouver C fractures occur distally remote from the implant. Vancouver B fractures are located at the level of the implant, and are subdivided into three types: B1 are fractures with a stable implant; B2 are fractures with a loose implant; and B3 are fractures with a loose implant in the presence of severe loss of bone stock.

Patient-specific classification systems

The Salter-Harris classification (24) was designed to classify pediatric fractures occurring around the growth plate of almost any bone. It consists of five types of fracture patterns **(Figure 2)**, ordered by increasing severity of the injury sustained by the growth plate. Type I is a complete separation of the epiphysis from the metaphysis without any fracture. In Type II, the fracture line extends along the epiphyseal plate to a variable distance and then goes through a portion of the metaphysis, thus producing a typical triangular-shaped metaphyseal fragment. Type III is an articular fracture, where the line of cleavage extends from the joint surface to the weak zone of the epiphyseal plate, and then extends along the plate to sever it from the metaphysis. Type IV is also articular, but here the fracture line extends from the joint surface of the epiphysis, across the full thickness of the epiphyseal plate and through a portion of the metaphysis, thereby producing a complete split. Type V is a crush injury of the growth plate; as no clear fracture line is visible, it is often discovered late when growth disturbances occur.

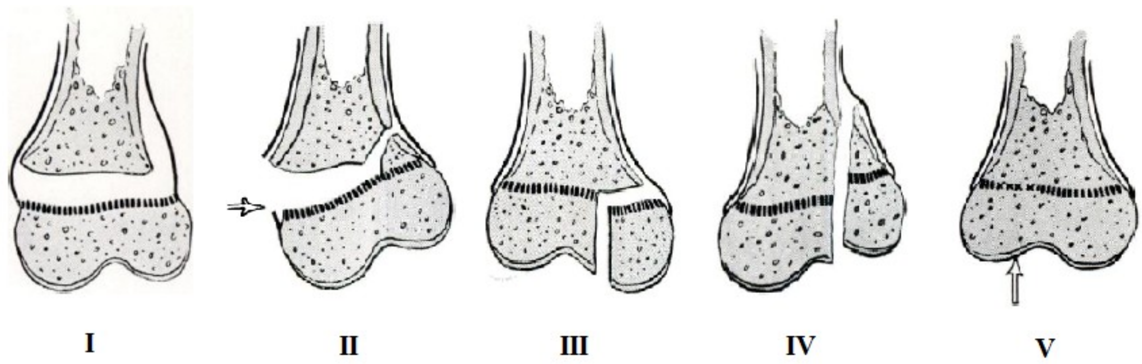

Figure 2 : Salter-Harris classification
Source : Salter RB, Harris WR. Injuries involving the epiphyseal plate. *J Bone Joint Surg (Am)* 1963; 45:587-622.

The classification gives some prognostic indications according to the type of fracture. Types I and II typically have a good prognosis as the growing cells that remain with the epiphysis are not injured. Type III, and even more so Type IV, may present with growth disturbances if the reduction is not perfect, since the growing cells may be damaged by the fracture line passing through them. The worst prognosis is associated with the Type V in which the growing cells are destroyed by the crush mechanism.

Mirels' scoring system (25) is a classification system of metastatic lesions of long bones. It was designed to predict the risk of pathologic fracture in any long bone, and thus evaluates the necessity for prophylactic surgery. It is therefore a kind of prophylactic classification system for pathologic fractures. Four characteristics of a metastasis are evaluated, each one receiving a score between one and three **(Figure 3)**. The final score is therefore a sum between 4 and 12. Lesions scoring ≤7 are at low percentage risk for fracture, so nonoperative treatment is advocated for those patients. For scores of >8 the risk of fracture is considered sufficiently elevated to advocate prophylactic fixation prior to other treatment such as irradiation. This is a good example of a classification system that has direct implication on the treatment of the patient, especially since it concerns prophylactic surgery. The difficulty with this scoring system for the observer is that the features of a metastasis can be fairly subjective: the patient's pain must be given a score, and the nature of the lesion (lytic, blastic or mixed) must be evaluated on a plain radiograph.

Variable	Score 1	Score 2	Score 3
Site	Upper limb	Lower limb	Peritrochanter
Pain	Mild	Moderate	Functional
Lesion	Blastic	Mixed	Lytic
Size	<1/3	1/3–2/3	>2/3

Figure 3 : Mirels' scoring system for impending pathological fractures
Source : El-Husseiny M, Coleman N. Inter- and intra-observer variation in classification systems for impending fractures of bone metastases. *Skeletal Radiol* 2010;39(2):155-160.

Generic or Universal system

In order to further investigation, evaluation, learning and teaching, the Arbeitsgemeinschaft für Osteosynthesefragen (AO) group sought from its early days to document the fracture cases treated by their members. The sheer amount of information, and the countless uncoordinated fracture classification systems that existed, made them realize that a universally applicable and acceptable system was needed. The development of such a system, the special project of Professor Maurice Müller, took many years and the classification system was finally published in 1990, with subsequent modifications and additions (3,26). The AO comprehensive classification of fractures of the long bones is the only generic or universal system in use today. *Universal* means that the same descriptive coding system of this classification can be applied to any bone of the skeleton. The AO classification of a fracture is a methodological process in five steps, corresponding to the answer to five questions:

Which bone? Numbers are attributed to the bones or limbs. 1 is for the humerus, 2 is for the forearm, 3 is for the femur, 4 is for the leg, etc. **(Figure 4)**

Which bone segment? Each bone has three segments: the proximal segment is coded as 1, the diaphyseal segment is 2, and the distal segment is numbered 3. The proximal and the distal segments of long bones are defined by a square whose sides are the same length as the widest part of the epiphysis. An exception is the malleolar segment of the ankle which received the special code 4.

Which fracture type? Fractures can be of three types: A, B or C. In the diaphysis (bone segment 2) the type A are simple two-fragment fractures, the type B are wedge fractures with one or more intermediate fragments but with some contact between the main fragments after reduction, and the type C are complex fractures with more than one intermediate fragment and no contact between the main proximal and distal fragments after reduction. At the ends of the long bone (bone segments 1 and 3), the type A are extra-articular fractures, the type B are partial articular fractures where a part of the articular surface is preserved and remains in continuity with the diaphysis, and type C are complete articular fractures with complete disruption of the articular surface from the diaphysis.

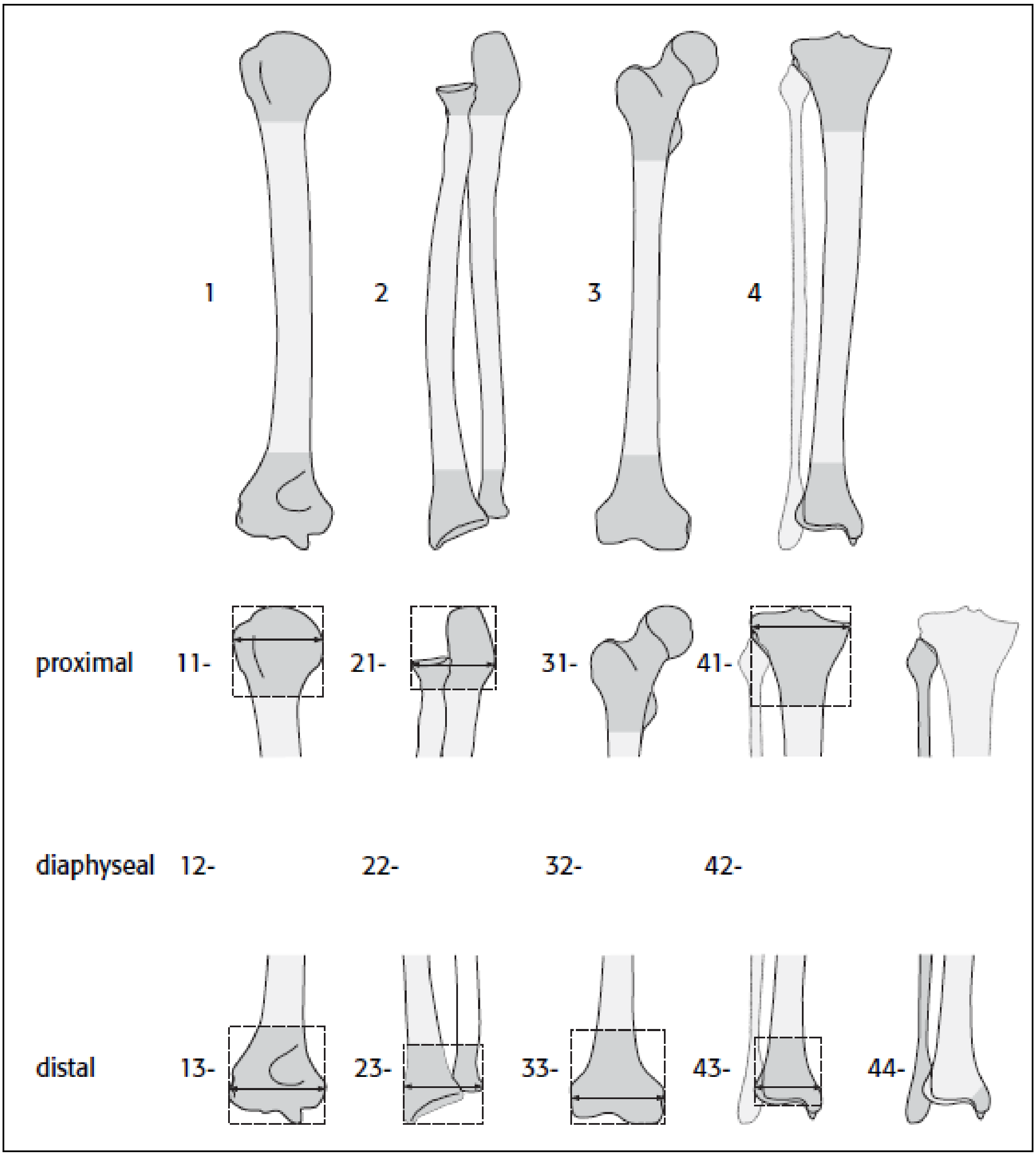

Figure 4 : AO Numbering of bones and bone segments. The first number designates the bone (the radius-ulna and the tibia-fibula are considered as one bone). The second number designates the segment: 1 for proximal, 2 for diaphyseal and 3 for distal. The malleolar segment is an exception, noted 44-. The proximal and distal segments are defined by a square whose sides are the same length as the widest part of the epiphysis (exceptions: 31- and 44-).
Source: Kellam JF, Audigé L. Fracture Classification. In: Rüedi TP, Buckley RE, Moran CG, eds. *AO Principles of Fracture Management*. Switzerland, AO Publishing;2007:69-85. Copyright by AO Foundation, Switzerland.

Which group? Each type of fracture is then divided into three groups (coded 1, 2 and 3) according to relevant details of the fracture, such as the angle of the fracture line or the degree of comminution. However, the definition of each fracture group is not constant and varies with the fracture types.

Which sub-group? Finally, each group of a fracture is divided into subgroups (coded .1, .2, and .3) according to key features of the fracture, in order to get the most precise description possible.

The result is a five-element alphanumerical code for the fracture. For example: a distal humeral fracture, complete articular and multifragmentary joint surface, and metaphyseal complex would be coded 13-C3.3.

As can be seen in **Figure 5**, for each bone segment there are 27 possibilities of fracture classification at the subgroup level. The fracture types, groups and subgroups are arranged in a theoretical ascending order of severity of the fracture. For example, a C fracture is theoretically more severe than an A fracture, and a B1 fracture is less severe than a B2 fracture. The principle stated during the design of the AO Comprehensive Classification was that the maximum amount of details about a fracture would lead to more accurate description and classification, which would in turn lead to a better understanding of the "essence" of that fracture. The result could then be a guide to treatment, improve research capabilities, and provide a prognostic outcome of the treatment.

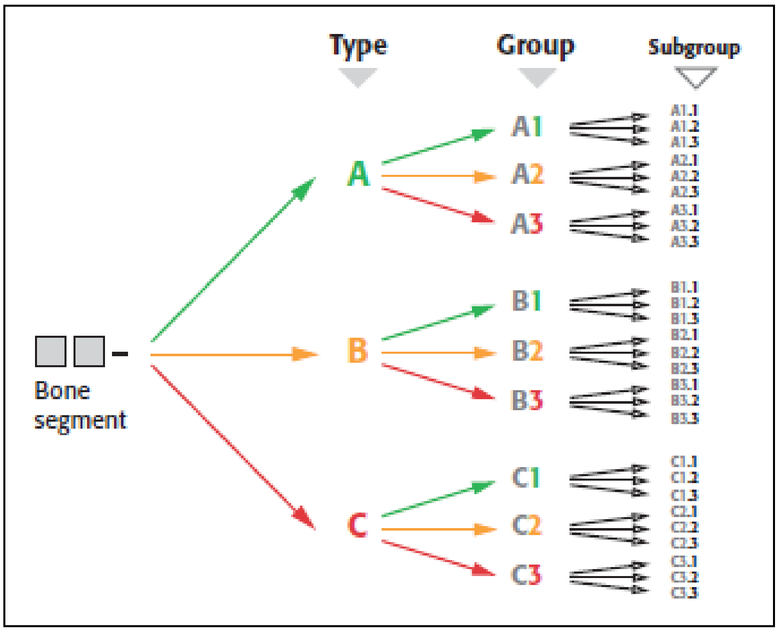

Figure 5 : Description of the morphology of the fracture, expressed in types, groups and subgroups according to the level of complexity of the description. The types, groups and subgroups are ordered in theoretical ascending severity according to the morphological complexity of the fracture, the expected difficulty of treatment and its prognosis.
Source: Kellam JF, Audigé L. Fracture Classification. In: Rüedi TP, Buckley RE, Moran CG, eds. AO Principles of Fracture Management. Switzerland, AO Publishing;2007:69-85. Copyright by AO Foundation, Switzerland.

The AO classification is still in constant evolution. It is continually evaluated by the AO group, changes are made according to evidence-based data, and new sections are developed. For example, a classification of long-bone fractures in children was published in 2006. The Orthopaedic Trauma Association (OTA) has adopted the AO classification system (27), and for most surgeons it is now known as the AO/OTA classification. Some journals, such as the *Journal of Orthopaedic Trauma*, have also decided to restrict the classification of fractures in their pages to the AO/OTA classification.

Soft-tissue injury classification systems

Fractures associated with soft-tissue injuries are much more complex to treat than low-energy fractures without soft-tissue concerns. For two equivalent fractures, the management protocol can be completely different in the setting of a soft-tissue injury. Typical fracture classification systems consider only the bone lesions and do not include the soft-tissues. Few authors have published classification systems for soft-tissue injury.

Gustilo and Anderson developed a classification system of open fractures on the basis of retrospective and prospective analysis of 1,025 patients (28). Further clinical experience led Gustilo to modify it to its present version (29), and this is the most widely used classification system of open fractures around the world. The classification integrates the severity of the fracture with the skin wound, the extent of soft tissue (muscles, periosteum and vascular elements) injury, and the degree of contamination. Gustilo type I fractures are associated with a clean wound of less than 1cm long. They are usually the result of a perforation from the inside out made by a sharp fracture fragment. Gustilo type II fractures are associated with a skin wound larger than 1cm, but without extensive soft tissue contusion, periosteal stripping, necrosis, flaps or avulsion. Gustilo type III are either open segmental fractures or fractures associated with extensive soft tissue damage, with or without gross contamination. They are subdivided in three types: type III-A are usually high-energy injuries, with extensive soft-tissue damage but with still enough soft-tissue coverage of the fractured bone. Type III-B fractures present massive soft-tissue loss, periosteal stripping and bone exposure, with no possibility of coverage. Type III-C defines any open fracture with associated vascular injury that requires repair.

Until recently, the only published classification of soft-tissue injury associated with closed fractures was the one described by Tscherne (30). There are four grades of increasing severity, as follows: Grade 0 are simple fractures with no or minimal soft-tissue injury, usually resulting from an indirect low-energy mechanism. Grade 1 includes fractures of mild to medium severity with superficial contusions or abrasions. The soft-tissue damage usually occurs through pressure from a bone fragment on the soft tissues from the inside. The Grade 2 fractures present deep contaminated abrasions with localized skin or muscle contusion. They usually result from a direct blow (such as a car "bumper" injury) causing a medium to severe fracture pattern. A fracture with an impending compartment syndrome is also part of Grade 2. The hallmarks of Grade 3 injuries are extensive skin contusion or crush, severe muscle damage and subcutaneous tissue avulsion. The fracture is usually severe and multifragmentary. Overt compartment syndrome or vascular injuries also belong to Grade 3.

Knowing quite well the major importance of soft-tissue injuries associated with fractures, the AO group developed a soft-tissue grading system with alphanumerical codes which complete their Comprehensive Classification of fractures (31). Three characteristics of soft-tissue injury are described: skin (fracture is either closed or open), muscle/tendon damage, and neurovascular injury. Each one is coded separately in increasing order of severity **(Figure 6)**. Skin lesions in closed fractures are coded as "IC" 1 to 5; skin lesions in open fractures are noted as "IO" 1 to 4; muscle/tendon injury is classified as "MT" 1 to 5; and neurovascular injury is coded as "NV" 1 to 5. The final result is a three-element code that is added to the AO classification of the fracture under consideration. For example, a closed simple transverse midshaft tibial fracture with the fibula fractured at the same level, associated with skin contusion, a one-compartment muscle injury, and no neurovascular injury would be classified as 42-A3.3 / IC2-MT2-NV1.

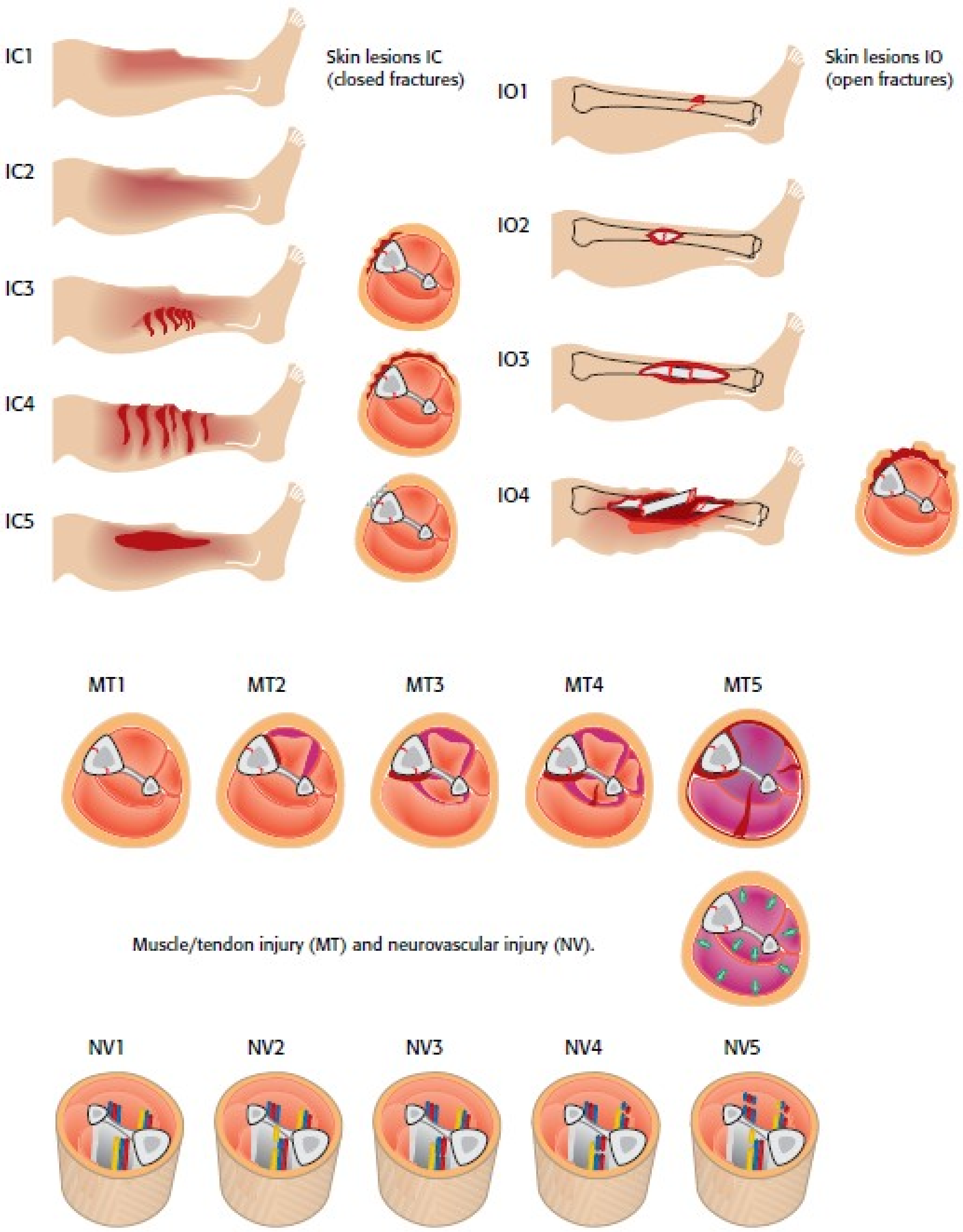

Figure 6 : The soft-tissue classification of the AO.
Source: Südkamp NP. The AO soft-tissue grading system. In: Rüedi TP, Buckley RE, Moran CG, eds. *AO Principles of Fracture Management*. Switzerland, AO Publishing;2007:99-112. Copyright by AO Foundation, Switzerland.

Skin lesions IC (closed fractures)	
IC 1	No skin lesion
IC 2	No skin laceration, but contusion
IC 3	Circumscribed degloving
IC 4	Extensive, closed degloving
IC 5	Necrosis from contusion

Muscle/tendon injury (MT)	
MT 1	No muscle injury
MT 2	Circumscribed muscle injury, one compartment only
MT 3	Considerable muscle injury, two compartments
MT 4	Muscle defect, tendon laceration, extensive muscle contusion
MT 5	Compartment syndrome/crush syndrome with wide injury zone

Skin lesions IO (open fractures)	
IO 1	Skin breakage from inside out
IO 2	Skin breakage from outside in < 5 cm, contused edges
IO 3	Skin breakage from outside in > 5 cm, increased contusion, devitalized edges
IO 4	Considerable, full-thickness contusion, abrasion, extensive open degloving, skin loss

Neurovascular injury (NV)	
NV 1	No neurovascular injury
NV 2	Isolated nerve injury
NV 3	Localized vascular injury
NV 4	Extensive segmental vascular injury
NV 5	Combined neurovascular injury, including subtotal or even total amputation

However, as useful as these classifications are, some surgeons think that the increasing complexity of these classification systems only shows that each major injury involving the soft tissues has its own and unique personality. The classification systems are useful for documentation, but in the setting of the emergent treatment demanded by fractures associated with soft-tissue injury the initial task for the surgeon is not to classify, but to describe the lesions as well as possible. Moreover the assessment of soft-tissue damage is a dynamic process that requires frequent re-evaluation because of the evolutionary nature of these injuries in the first days of management following the fracture (32).

Fracture classification systems: characteristics and statistics

In the December, 1993 edition of the *Journal of Bone and Joint Surgery (American)*, Dr. Albert H. Burstein wrote a famous editorial about fracture classification systems. As he said, fracture classification systems are, indeed, tools. The main purpose of these tools is to help the orthopaedic surgeon choose an appropriate method of treatment for each and every fracture in their practice. They should also provide a reasonably precise estimation of the outcome of that treatment (19). Other authors have also added that classification systems should facilitate and clarify communication between physicians, and assist the documentation, research and comparison of published results (33-36).

However, we all like our tools to be of the finest quality. That is, they should do the proper work they were designed for with constancy over time, and that we can trust the information they give us. In order to be the best of tools, the ideal classification system must have seven qualities: validity, reliability (reproducibility), repeatability, all-inclusiveness, mutual exclusiveness, logic, and clinical usefulness (19,33,34,37).

Validity is the capacity of the system to precisely describe the true state of the pathologic process. It is the correlation between the classification system and the reality. To quantify its validity, the tool in question must be compared to some *gold standard*, as the true reality is made unreachable by the unavoidable observer interpretation. In a system based on fracture lines, a gold standard could be an intraoperative assessment of the fracture lines. The problem in our era of indirect reduction and percutaneous fixation or minimally invasive approaches is that finding and measuring a gold standard is usually very difficult, and thus also the measure of validity.

Reliability (also referred to as reproducibility) is defined by the fact that a given fracture is classified as the same by several observers. This is known as the interobserver agreement. However, reliability is not synonymous with validity. For example, if an ankle fracture is classified as Weber B by all observers, the measurement is reliable. But if the intraoperative findings are of a Weber C fracture, then the classification was not valid.

The repeatability of a classification implies that the same observer classifies a given fracture always the same on several different occasions. This is known as the intraobserver agreement.

All-inclusiveness and mutual exclusiveness means that every possible fracture of a given anatomical region must fit one and only one category of the related fracture classification. In a study by Maripuri, a high number of proximal tibia fractures were unclassifiable with the Hohl and Moore system, thereby showing that this classification system is not all-inclusive (38).

Logic and clinical usefulness are self-defined. A classification system which is not logical is only a source of confusion, misinterpretation, and misuse. And a classification system that is not useful in everyday orthopaedic clinical practice is of no use at all, and should be quickly forgotten. Some authors even think that the clinical issue of agreeing on a treatment plan is the most important goal of a fracture classification system (39).

Because of the difficulty of measuring validity, a fracture classification system should have at least high degrees of reliability and repeatability (19,34). The basic method of evaluating these two parameters is to measure the raw observed agreement, expressed as the percentage of times that different observers agreed on their assessments. For example, if observers agreed on 75 of 100 assessments, the observed proportion of agreement would be 0.75, or 75%. The limitation of this method is that the chance factor is not taken into account, so there is no indication as to what part of the observers' agreement is made by chance only. To assess agreement that occurred above and beyond that related to chance alone, Cohen introduced the Kappa statistic (κ) in 1960 (40). The Kappa statistic provides a pair-wise proportion of agreement between observers corrected for chance. It is expressed mathematically as the observed agreement minus expected chance agreement, divided by the maximum possible agreement not related to chance **(Figure 7)**. The expected chance agreement is the percentage of agreement attributed to chance alone. It is a statistical calculation that is dependent upon the number of observers, the number of choices in each assessment, and the number of assessments (41).

		Observer 1		
		A	**B**	**total**
Observer 2	**A**	50	20	70
	B	10	20	30
	total	60	40	100

Observed agreement: $\frac{50+20}{100} = 0.7$

Chance agreement: $\frac{\text{expected agreement for A + expected agreement for B}}{\text{total}} = \frac{\frac{60\times 70}{100}+\frac{30\times 40}{100}}{100} = 0.54$

Agreement beyond chance: $\kappa = \frac{\text{observed agreement} - \text{chance agreement}}{\text{1-chance agreement}} = \frac{0.7-0.54}{1-0.54} = 0.34$

Figure 7 : Calculation of the Kappa statistic
Example with the simplest situation: two observers doing 100 observations with a clear cut "A" or "B" choice (which could be "fracture" or "no fracture", for example). The observed agreement is the total of the observations on which both observers agreed, divided by total number of observations (in this case agreement of 0.7, or 70%). Chance agreement is calculated. Then the κ statistic can be calculated, expressing the agreement between both observers beyond chance only.

The κ value was designed to analyze categorical data, which have to be divided into two types. *Nominal* (unranked) data are given equal importance between all categorical differences. For example, there is equal importance between blue and brown eyes, and between blue and green eyes; none is better or worse than the others. In *ordinal* (ranked) data, the difference between some categories is given more credit, or more "weight", than the difference between some other categories. For example, the difference between Gustilo 1 and Gustilo 2 open fractures is less important than between Gustilo 1 and Gustilo 3 open fractures, as they are ranked according to an increase in severity. To analyze ordinal data, Fleiss introduced the weighted κ statistic (42), in which a "weight modifier" gives some credit to partial agreement in order to reflect the inequality between the different categories. While unweighted κ must always be used for nominal data, a choice must be made whether or not to use weighted κ for ordinal data (34). Weighting κ values gives rise to two problems. As the chosen weights can greatly alter the κ values, it is mandatory that the weighting plan is clearly defined in advance. And without uniform weighting schemes, no comparison between studies is possible if the used scheme is not precisely described.

Although most authors have now accepted the κ value as method to measure observer agreement, its interpretation is somewhat difficult. Values obtained range from -1 to +1, where -1 corresponds to perfect disagreement, +1 is a perfect agreement, and 0 corresponds to agreement due to chance only. Between these reference values, no statistically defined cut-off values express the level of agreement between observers, even if it seems logical that the higher the value, the more reliable the classification system. The two most widely used scales of κ level of agreement published are those of Landis and Koch (43) and Svanholm (44) **(Figure 8)**. Despite their widespread acceptance and use, it is important to note that the cut-off values of these two guidelines were chosen arbitrarily. It is also important to note that κ is dependant both on the number of categories (i.e., its value is greater if there are less categories) and the prevalence of the condition (45). For example, a category with a high prevalence can give rise to a high raw interobserver agreement but with a very low κ. Therefore care must be taken when comparing κ from different studies, and the prevalence of the analyzed conditions (fracture categories) should be clearly indicated, which is in fact rarely done.

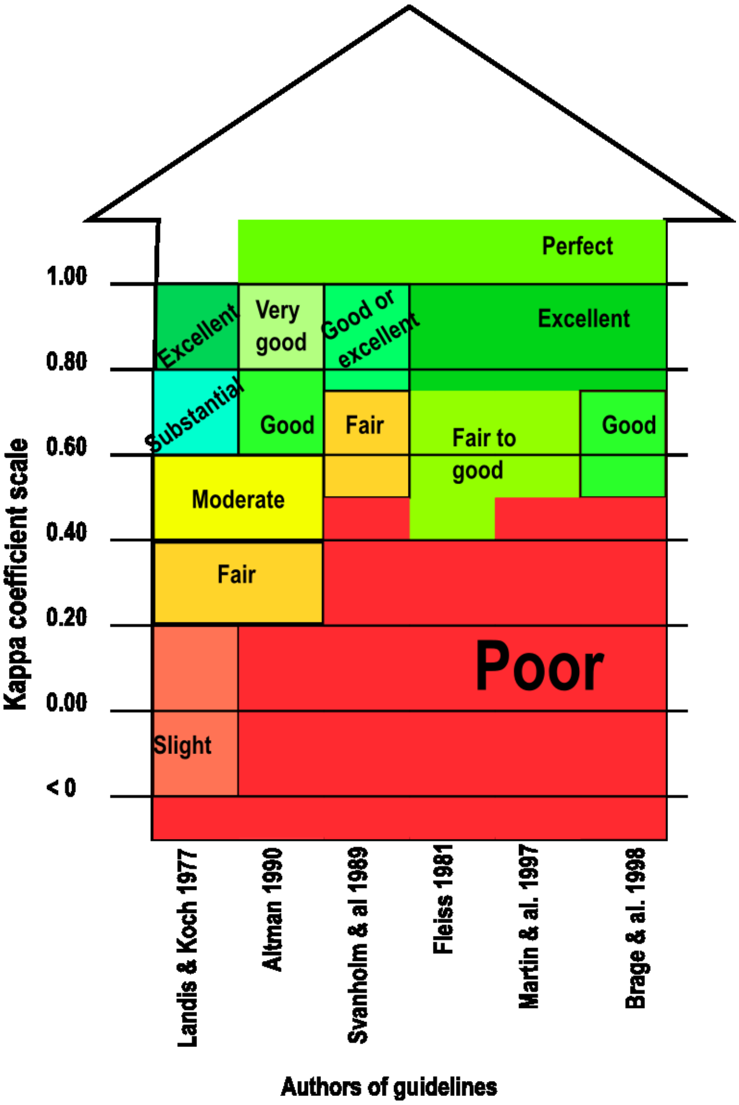

Figure 8 : Guidelines used for the interpretation of the Kappa coefficient

In the case of numerical variables (data falling on a continuum; for example, the amount of displacement in millimeters), an index of reliability commonly used to measure reproducibility and repeatability is the Intraclass Correlation Coefficient (ICC), where values range from 0 (no agreement) to 1 (perfect agreement) (45). However the ICC is of limited use as it is not related to the size of the error which is clinically acceptable, and its values shouldn't be compared between different sets of data as the ICC is influenced by features of the data (for example the ICC will be higher if the observations are more variable).

Limitations and flaws of current fracture classification systems

Since Burstein's editorial, many authors have evaluated the functionality of the most common fracture classification systems in terms of intraobserver and interobserver reliability. Unfortunately the next step in the process proving the usefulness of these tools has not been achieved because most of the classification systems show disappointing reliability. One of the few fracture classification systems considered good and clinically useful on the basis of scientific testing is the Weber classification of ankle fractures. A study by Malek et al. (35) showed substantial interobserver (raw agreement of 78%, mean κ 0.61) and intraobserver (raw agreement of 85%, mean κ 0.74) reliability. A second one is the Vancouver classification of periprosthetic fractures, where the European validation study of Rayan et al. (46) showed substantial agreement for all observers with a maximum κ of 0.74. And a third one is the Letournel classification of acetabular fractures. Beaulé (47) demonstrated a substantial interobserver and intraobserver reliability, with κ values of 0.69 and 0.77, respectively.

Almost every other classification system tested shows unsatisfactory reliability, whether they are commonly or rarely used, simple or complex. Examples are numerous. In two studies, a total of five classification systems of distal radius fractures were reviewed by Ploegmakers et al. (48) and Belloti et al. (49). The interobserver agreement was unsatisfactory in all of them. They concluded that the use of the AO, Frykman, Fernandez, Older, and Cooney's Universal classifications cannot be recommended for clinical application because of their questionable reliability and reproducibility. Two classifications of radial head fractures were tested by Sheps et al. (50). The Hotchkiss modification of the Mason classification showed only moderate interobserver reliability, confirming the results of Morgan et al. (51). The interobserver reliability for the AO classification was only fair at the subgroup level (rising to moderate without subgroups), with the major concern that this classification was unable to differentiate fractures needing operative versus conservative treatment. As regards trochanteric fractures, neither the AO classification (52-54) nor the Jensen classification (52,54) met acceptable thresholds for reliability. van Embden (54) and Fung (52) also showed that surgeons were unable to reliably determine fracture stability or instability, thus raising concerns about previous studies that recommended implant choice on the basis of fracture stability. Primarily two classification systems of tibial plateau fractures have been examined. Walton et al. (36) found the AO classification to be slightly superior to the Schatzker classification in terms of interobserver reliability, but Maripuri et al. (38) found the opposite. However both authors agree that neither classification system is good since the interobserver reliability is at best moderate with a mean κ of 0.47. Where does the problem lie, and why are there so few reliable and repeatable fracture classification systems?

Classification system flaws

Some fracture classification systems have inherent flaws which make them inevitably unreliable. In their study of tibial plateau fracture classification systems, Maripuri et al. (38) showed that the Hohl and Moore system was not all-inclusive since many fractures were unclassifiable. When trying to classify fractures with this system an observer would then be forced to choose the least wrong category instead of the best one, thereby leading to imprecision and variability between observers.

Another common problem arises when the classification is based on non-radiological factors, such as bone quality, implant stability or fracture stability. These are clinical and dynamic factors which are very hard to evaluate on a static radiograph which is only a glimpse in the life of a fracture. Although the study of Rayan et al. (46) showed the validity of the Vancouver classification, the authors also agreed that there was difficulty in determining preoperatively the stability of the prostheses and the quality of the bone stock, and thus to establish the diagnosis of a B1- versus a B2-fracture. The repercussion is a lower inter-observer κ in the B sub-group (κ 0.67) compared to the whole inter-observer κ of 0.74.

The same issue arises with fractures in the elderly where the presence of osteoporosis could theoretically influence the pattern of the fracture, but the bone quality is not taken into account in the current classifications as noted by Hoesel (55). The main reasons for this is are that, first, there is still no consensus on which fractures can be considered as “osteoporosis-related” (56), and second, recent studies failed to find a correlation between the severity of the osteoporosis and the severity of fractures commonly considered as osteoporosis-related (pertrochanteric(57) and distal radius(58) fractures). It seems then that the bone quality has no influence over the severity of the fractures and so this factor probably doesn't bring any additional value to the current classification systems.

Some fractures are commonly evaluated on the basis of their assumed stability or instability, the most famous example being proximal femur fractures. Fracture stability is in fact very difficult to evaluate on the basis of static radiographs. As was previously mentioned, the study of Fung et al. (52) where 12 reviewers evaluated 56 radiographs of intertrochanteric fractures showed unacceptable reliability for both AO/OTA and Evans/Jensen classification. In addition, surgeons were unable to determine fracture stability when specifically asked to do so. The study of van Embden et al. (54) obtained the same results with 10 reviewers examining 50 trochanteric fractures, with the additional result that after considering the postoperative radiographs, the reviewers concluded that 15% to 18% of the fractures were treated with an inappropriate type of implant. Thus, basing a classification system and the choice of treatment on the aspect of fracture stability seems to be a mistake. In fact, a problem with evaluation of fracture stability is the complete lack in the literature of a clear and consensual definition of what is "stable" or "unstable." Interestingly, 11 observers classifying 34 subcapital hip fractures obtained only fair results with the Garden classification, but demonstrated almost perfect agreement between the most experienced of them when they classified the same fractures as "stable" or "unstable" according to the precise definition established by the study authors (59). Beimers et al. then concluded that the Garden classification is unreliable and should be abandoned in favor of categorizing these fractures as stable versus unstable. Probably when a clear consensus-based definition of fracture stability is published, new studies will determine if the concept of stability is effectively a source of imprecision and variability.

Experience of the observer

Some authors have postulated that the level of professional experience of the observer could be an important factor affecting the reliability of fracture classification (53). Experienced observers are supposed to be well informed about classification diagrams and accustomed to classifying fractures, and thus should be more accurate and less variable in their classification process than inexperienced observers. This theory has only rarely been proven, as with the Vancouver classification of periprosthetic femoral fractures (46) or with the Letournel classification of acetabular fractures (47). Actually, many studies about classification of long bone fractures have demonstrated that the experience of the observer has no influence on reliability, whether it be fractures of the proximal humerus (60), the distal radius (48,49), the proximal femur (54), the tibial plateau (36), or the distal tibia (61), or in the setting of the general AO classification (62).

Radiographic images: Their quality, the difficulties of identifying fracture lines, and the role of new technologies

The quality of fracture radiographs can vary because of a number of factors, including the type of radiographic machine, the skill of the radiologic technician, or the physical characteristics of the patient. It would seem logical that poor-quality radiographs may affect the observer's ability to accurately and reliably classify fractures, especially if the fracture lines are difficult to see.

However, a study specifically evaluating the impact of the radiographs' quality did not show it to be a significant source of interobserver variability (63). Actually, interobserver agreement on the adequacy of the radiographs was poorer than agreement on the classification of the fractures themselves. It therefore appears that improving the quality of plain radiographic images is unlikely to improve the reliability of classification of fractures, at least in the case of tibial plafond fractures.

Even with good-quality radiographs, it may be sometimes difficult to identify the fracture lines, notably in the context of articular fractures with multiple overlapping fragments, factors related to the complex three-dimensional (3D) shape of the bone, or with osteoporotic bone. In the above-cited study (63), Dirschl et al. asked observers to identify and mark the fragments of tibial pilon fractures before classifying them, which did not improve the interobserver reliability over that from a previous session without drawings. However, when the fragments were identified and marked beforehand by the senior author of the study, the interobserver reliability was significantly improved, but only to a moderate level of agreement (κ 0.54). These results show that the identification of fracture lines and fracture fragments is difficult. However, it is only one factor in interobserver variability since when this factor was removed by pre-marking the fragments, the agreement did not rise to an almost perfect level. Another study (41) confirmed these facts in the context of tibial plateau fractures, where the authors concluded that the reliability of fracture classification is limited by the observers' ability to agree on basic radiographic assessments such as the location of fracture lines or the amount of displacement and comminution.

Although most classification systems were designed on the basis of plain radiographs, the advent of new imaging technologies, especially CT, ushered in the hope of increasing classification reliability largely due to the improved detail and specific information that they could provide. Unfortunately many studies have shown that two-dimensional (2D) CT scans do not improve the intraoberver and interobserver agreement on fracture classification (39,41,47,61,64-66). Paradoxically, one explanation could be the increased difficulty for the observer to analyze the huge amount of information provided by CT with its multiple imaging planes, to follow the fragments from one image to another, and to imagine a complex 3D volume like an articular surface with 2D image slices (65,67). However, standard 2D CT still has advantages for the characterization of the fracture in terms of better quantification of articular surface incongruity (68,69), and also for increasing observers' agreement on treatment plan (39,70).

Recent studies have evaluated more sophisticated post-acquisition treatment of CT images with promising results. Harness et al. (67) found that in comparison to 2D CT, 3D reconstruction of CT scans improved both the reliability and the accuracy of radiographic characterization of articular fractures of the distal part of the radius as well as treatment decisions, but without knowledge whether this would have resulted in better patient outcomes or more cost-effective treatment. Hu et al. (71) showed that the interobserver reliability for both the AO and Schatzker tibial plateau classification systems improved from "substantial" with the use of plain radiographs combined with 2D CT images, to "almost perfect" with the use of plain radiographs and associated 3D CT images. In another study, Doornberg et al. (72) showed that 3D CT significantly improved both the intraobserver and the interobserver agreement for the characterization, classification, and treatment of distal humeral fractures, but only to a “moderate” maximal level of agreement. This led them to conclude that there is substantial disagreement among qualified observers that cannot be resolved even with more sophisticated imaging techniques. In an example of proximal humerus fractures, Sjödén et al. (73) showed unsatisfactory reliability of both the AO and Neer classification systems, with absolutely no improvement with 2D or 3D CT imaging. Pushing the technological sophistication even further with the use of 3D-volume rendering and special stereo-visualization workstations, Brunner et al. (60) improved both intraobserver and interobserver reliability of these classifications to “good” and even “excellent”. Thus, while it seems that in the present era of high-definition video and three-dimensional cinema these new radiographic imaging technologies could help us to better evaluate and classify fractures, the questions remain as to the increased cost and time for such imaging studies.

Complexity of fracture classification systems

“To make a fracture classification work, one must keep it simple,” said Dr Sanders (74). It could effectively seem reasonable that a too complex classification system would lead the observer either to difficulties in understanding the basic system or to hesitation between too many categories, thus resulting in uncertainty and variability of classification. For example, in the AO fracture classification system, a given fracture can belong to one of 3 types, or one of 9 groups, or one of 27 subgroups, depending upon the amount of detail of the description. Several studies have shown that observers’ reliability drops with every increase in the classification’s complexity (from “type” to “group,” and from “group” to “subgroup”) (36,53,61,75), leading to the conclusion that acceptable reliability was only achieved at the “type” level. Thus, drawing guidelines concerning fracture management based on patterns more complex than the broad AO type fracture classification was meaningless (74).

However, the complexity of the AO classification is probably not the only factor that explains its poor performance at higher levels of detail description. Two studies tried to improve the reliability of the AO classification within the setting of tibial pilon fractures (63) and ankle fractures (76) by simplifying the diagnostic process with a binary decision-making protocol. When evaluating the radiographs, the observers could not jump directly to the diagnosis but were forced to follow a path of thought by answering sequential binary questions (whose answers could be only “yes” or “no”) which led them to the final classification code. In both studies there was no statistically significant difference in reliability between the original and binary classification systems. Moreover, apparently more complex systems like the Letournel classification for acetabular fractures perform better in terms of reliability than simpler classification systems (47). The simplification of the classification process or the application of binary decision-making does not appear to be effective in improving interobserver reliability in fracture classification.

Simplicity still has its advantages in everyday clinical practice notably for ease in communication, but lacks the level of detail necessary for research purposes which is in turn too cumbersome to be used in clinical practice. Therefore Bernstein advocated that two classification systems be used for every fracture: one simple and succinct for clinical use and another detailed enough for research purposes (77). Such a dual system would probably add to the confusion rather than solve the problem of reliability (78), which is why Colton advocated the use of a multi-layer system, with increasing details about the personality of the fracture as the classifier descends through it, and with an upper layer serving as an everyday working tool for the surgeon. A compromise like this would not be a weakness, but a foundation on which to build.

Reliability of reliability studies

Although some fracture classification systems appear better than others, and some should no longer be used because of their evident flaws, care should be taken before recommending or discarding a classification system. While a number of studies have been conducted to evaluate the reliability of many fracture classification systems, the quality of these studies was not always optimal. Audigé et al. (79) reviewed 44 studies assessing 32 fracture classification systems and found considerable variation in their methodologies. For example, in these 44 studies, the study population was clearly defined by inclusion/exclusion criteria in only 59%, the selection of cases was considered representative of the study population in only 39%, not a single study justified the size of the sample chosen, the participating raters were judged representative of the eventual users of the classification in only 9%, and the number of raters was appropriate in 23%. The statistical analyses, with the Kappa coefficient used in 88% of the studies, seemed adequate for the study objectives in only 39% of the studies. Audigé et al. thus recommended the results and conclusions of these reliability studies should be interpreted in the light of their methodological strength. They also pointed out the need for methodological standards for reliability studies.

Current usefulness and qualities of fracture classification systems

Classification systems have four purposes: naming things, grouping objects of the same category, predicting outcomes and guiding actions. Almost every possible fracture fits into at least one classification, and thus has a name that is usually an eponym. The usefulness for communication between orthopaedic surgeons in everyday practice is obvious, for even without a radiograph one can have quite a good mental image of a fracture only by its name. However the classification must be well-known by both users, and some communication problems and confusion could arise with the multiplicity of classification eponyms for the same fracture.

More than easing communication and passing on knowledge to trainees, classification systems also have an educational role. In order to correctly classify a fracture, the bony anatomy must be well-known, the mechanism of injury must be understood, and the different characteristics of the fracture itself must be established, which implies a certain discipline of thought. Nonetheless, we agree with Smith (80) that "in practice, most surgeons classify occasionally (on courses), a few formally and even fewer have protocols which plan management around a classification system. For the most part we continue to describe a fracture in 'longhand' with regard to its site, pattern, displacement and complicating features." Even if this *longhand* method probably gives rise to inaccuracies, many formal classification systems cannot claim to be more accurate or reliable.

How we do it

At our institution we use the *longhand* method of fractures description for daily communication, mixed with the commonly used eponymic classifications that our masters taught us, like Garden, Schatzker or Weber. They have the advantages of being simple, known to all, and very quickly descriptive of the situation. It is also very interesting to note how psychologically difficult it is not to use what has been taught to us since our first years of internship, even when we know it's not optimally accurate or reliable. The inaccuracies that our *longhand* method could induce are balanced by the immediate availability of the radiographic images that are always reviewed by a senior before any therapeutic decision is made.

However, for research purposes, all patients hospitalised or operated on must have their fracture(s) classified according to the AO classification system. The therapeutic decision is not made on the basis of this classification. This coding allows us to easily retrieve specific injuries for studies and to compare our data with those of the literature.

Because of their limitations in interobserver and intraobserver reliability, current classification systems probably fail in their last two purposes, predicting outcomes and guiding actions. This substantial variability casts doubt on comparative studies that have inferred a best treatment (or implant) choice on the basis of the fracture classification. In the same way, there is nothing in orthopaedic literature to date that validates an outcome prognosis with a fracture classification. Only one study (18) of isolated unilateral lower limb fractures was specifically designed to evaluate the outcome prognosis of the AO classification, and found no correlation between the alphanumerical code and six to twelve month functional performance and residual impairment.

Considerations for the future

There is still much work to do to find the optimal fracture classification system which will reliably guide the orthopaedic surgeon's decision as to best treatment, and as well predict outcomes. Precise classification is mandatory in order to improve patient care, as well as for hospital managers and administrators who need accurate information to recover the appropriate costs for treatment. Improvements in the concept of fracture classification have to be global and not only centered on bone radiography. Before deciding upon a course of treatment, one must be aware of the many variables that that make up the personality of the injury and need to be included in any classification system.

These are not only the specific musculoskeletal trauma involving bone, cartilage and soft tissue. They also include the patient and factors such as age, occupation, medical condition, needs, expectations, motivation, psychological status, level of education and socioeconomic status.

New imaging technologies like 3D CT volume-rendering with stereo visualization will probably help us to better understand and categorize fractures, hopefully in a more reliable way. Improved complementary imaging modalities with MRI and ultrasound should further provide us with more information about possible soft-tissue and cartilage injuries, and the health of the bone at the fracture site. The data processing of the images is also progressing to the point where it can provide a huge amount of new or improved information with the same raw data as that in the past. One could even envisage the development of computer algorithms that could automatically recognize the fracture patterns on digital images and classify them according to a selected classification system, in the same way that today computers read electrocardiograms. However, one must also consider important aspects in the use of these new technologies that are safety, expenses and accessibility (81). Newest technologies are more expensive and not available in every hospital, and the increasing use of certain modalities like CT or contrast injection could raise some concerns about patient safety.

Improvements could also be made by combining statistical methods and information technologies, as Bernstein demonstrated (82). The concept of “Crowd intelligence” is to aggregate information from a group of people to get a consensus that yield better decisions or more precise estimations. Applied to fracture classifications, the radiographs were sent to several readers by e-mail and the answers gathered in real-time. That way, a virtual group of 5 orthopaedic surgeons obtained a K of 0.8 for inter-observer reliability for the diagnosis of displacement, in less than three hours. However, increasing the number of participants in the consensus group only enhances the proclivities of each individual. So if individual accuracy is less than 50%, the crowd intelligence would indeed decrease accuracy. For classifications with a decent basic reliability, one could imagine a mobile application that would make use of crowd intelligence in order to improve reliability.

Nonetheless some characterizations about the basic features of fractures are still missing. Experts should work together and set precise definitions of, for example, fracture stability, displacement or comminution. They should also agree on precise methodologies for studies that evaluate the classification systems, and define the statistical cut-off of what is an acceptable or unacceptable reliability. In addition to the technical feats that show promising results but also limitations, there are two ways of improving fracture classification. Existing systems can be modified, or brand-new systems can be created. Audigé et al. (33) defined a precise method to modify or create a fracture classification system in the most efficient way, which includes three phases of validation: a pilot phase based on expert consensus; a second phase of multicenter consensus; and a third phase of prospective clinical study. A new classification of paediatric long bone fractures has since been developed that way (83) and has shown a good overall interobserver agreement of K=0.71.

Although time- and resource-consuming, this rigorous method is probably the best way to achieve the creation of valid, useful and reliable classification tools for our orthopaedic practice.

The classification of long bone fractures is currently undergoing a revolution. What will emerge from this will probably change completely our practice habits, but will also greatly advance our understanding of specific fractures and thus improve the quality of care we render our patients.

References

1. **Linnaeus C.** Systema naturae in quo naturae regna tria, secundum classes, ordines, genera, species, systematice proponuntur. Stockholmiae: Kiesewetter Gottfried; 1740.
2. **Woese CR, Kandler O, Wheelis ML.** Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990; 87(12): 4576-9.
3. **Müller ME, Nazarian S, Koch P.** The Comprehensive Classification of Fractures of Long Bones. Berlin, Heidelberg, New York: Springer-Verlag; 1990.
4. **Breasted JH.** The Edwin Smith Surgical Papyrus: Hieroglyphic Transliteration, Translation And Commentary V1. Whitefish: Kessinger Publishing; 2006.
5. **Hippocrates.** On Fractures. Whitefish: Kessinger Publishing; 2004.
6. **Celsus.** Cure générale de la fracture du bras, de l'avant-bras, de la cuisse, de la jambe et des doigts. In *Traité de médecine de Celse, d'après l'édition de Léonard Targa*, Paris: Imprimerie de Béthune et Plon; 1838.
7. **Pott P.** Some few general remarks on fractures and dislocations. London; 1765.
8. **Peltier LF.** Eponymic fractures: Giovanni Battista Monteggia and Monteggia's fracture. Surgery. 1957; 42(3): 585-91.
9. **Colles A.** On the fracture of the carpal extremity of the radius. Edinb Med Surg J. 1814(10): 182-6.
10. **Schepers T, van Lieshout EM, Ginai AZ, Mulder PG, Heetveld MJ, Patka P.** Calcaneal fracture classification: a comparative study. J Foot Ankle Surg. 2009; 48(2): 156-62.
11. **Garden RS.** Low angle fixation in fractures of the femoral neck. J Bone Joint Surg Br. 1961(43): 647-663.
12. **Neer CS, 2nd.** Displaced proximal humeral fractures. I. Classification and evaluation. J Bone Joint Surg Am. 1970; 52(6): 1077-89.
13. **Sanders R, Fortin P, DiPasquale T, Walling A.** Operative treatment in 120 displaced intraarticular calcaneal fractures. Results using a prognostic computed tomography scan classification. Clin Orthop Relat Res. 1993(290): 87-95.
14. **Zwipp H, Tscherne H, Wulker N, Grote R.** [Intra-articular fracture of the calcaneus. Classification, assessment and surgical procedures]. Unfallchirurg. 1989; 92(3): 117-29.
15. **Dirschl DR, Dawson PA.** Injury severity assessment in tibial plateau fractures. Clin Orthop Relat Res. 2004(423): 85-92.
16. **Marsh JL, Buckwalter J, Gelberman R, Dirschl D, Olson S, Brown T, Llinias A.** Articular fractures: does an anatomic reduction really change the result? J Bone Joint Surg Am. 2002; 84-A(7): 1259-71.
17. **Oakes DA, Jackson KR, Davies MR, Ehrhart KM, Zohman GL, Koval KJ, Lieberman JR.** The impact of the garden classification on proposed operative treatment. Clin Orthop Relat Res. 2003(409): 232-40.
18. **Swiontkowski MF, Agel J, McAndrew MP, Burgess AR, MacKenzie EJ.** Outcome validation of the AO/OTA fracture classification system. J Orthop Trauma. 2000; 14(8): 534-41.
19. **Burstein AH.** Fracture classification systems: do they work and are they useful? J Bone Joint Surg Am. 1993; 75(12): 1743-4.
20. **Schatzker J.** Compression in the surgical treatment of fractures of the tibia. Clin Orthop Relat Res. 1974(105): 220-39.
21. **Schatzker J, McBroom R, Bruce D.** The tibial plateau fracture. The Toronto experience 1968--1975. Clin Orthop Relat Res. 1979(138): 94-104.
22. **Weber BG.** Die Verletzungen des oberen sprunggelenkes. Bern; 1972.

23. **Duncan CP, Masri BA.** Fractures of the femur after hip replacement. Instr Course Lect. 1995; 44: 293-304.
24. **Salter RB, Harris WR.** Injuries involving the epiphyseal plate. J Bone Joint Surg Am. 1963(45): 587-622.
25. **Mirels H.** Metastatic disease in long bones. A proposed scoring system for diagnosing impending pathologic fractures. Clin Orthop Relat Res. 1989(249): 256-64.
26. **Kellam JF, Audigé L.** Fracture Classification. In *AO Principles of Fracture Management*, pp. 69-85. Rüedi TP, Buckley RE, Moran CG, Switzerland: AO Publishing; 2007.
27. Fracture and dislocation compendium. Orthopaedic Trauma Association Committee for Coding and Classification. J Orthop Trauma. 1996; 10 Suppl 1: v-ix, 1-154.
28. **Gustilo RB, Anderson JT.** Prevention of infection in the treatment of one thousand and twenty-five open fractures of long bones: retrospective and prospective analyses. J Bone Joint Surg Am. 1976; 58(4): 453-8.
29. **Gustilo RB, Mendoza RM, Williams DN.** Problems in the management of type III (severe) open fractures: a new classification of type III open fractures. J Trauma. 1984; 24(8): 742-6.
30. **Tscherne H, Gotzen L.** Fractures With Soft Tissue Injuries. Berlin, Heidelberg, New York: Springer-Verlag; 1984.
31. **Südkamp NP.** The AO soft-tissue grading system. In *AO Principles of Fracture Management*, pp. 99-112. Rüedi TP, Buckley RE, Moran CG, Switzerland: AO Publishing; 2009.
32. **Masquelet AC, de Haas W.** The problem of classifying soft-tissue lesions. In *AO Principles of Fracture Management*, pp. 373-374. Rüedi TP, Buckley RE, Moran CG, Switzerland: AO Publishing; 2007.
33. **Audige L, Bhandari M, Hanson B, Kellam J.** A concept for the validation of fracture classifications. J Orthop Trauma. 2005; 19(6): 401-6.
34. **Garbuz DS, Masri BA, Esdaile J, Duncan CP.** Classification systems in orthopaedics. J Am Acad Orthop Surg. 2002; 10(4): 290-7.
35. **Malek IA, Machani B, Mevcha AM, Hyder NH.** Inter-observer reliability and intra-observer reproducibility of the Weber classification of ankle fractures. J Bone Joint Surg Br. 2006; 88(9): 1204-6.
36. **Walton NP, Harish S, Roberts C, Blundell C.** AO or Schatzker? How reliable is classification of tibial plateau fractures? Arch Orthop Trauma Surg. 2003; 123(8): 396-8.
37. **Martin JS, Marsh JL.** Current classification of fractures. Rationale and utility. Radiol Clin North Am. 1997; 35(3): 491-506.
38. **Maripuri SN, Rao P, Manoj-Thomas A, Mohanty K.** The classification systems for tibial plateau fractures: how reliable are they? Injury. 2008; 39(10): 1216-21.
39. **Chan PS, Klimkiewicz JJ, Luchetti WT, Esterhai JL, Kneeland JB, Dalinka MK, Heppenstall RB.** Impact of CT scan on treatment plan and fracture classification of tibial plateau fractures. J Orthop Trauma. 1997; 11(7): 484-9.
40. **Cohen J.** A coefficient of agreement for nominal scales. Educational and Psychological Measurement 1960(20): 37-46.
41. **Martin J, Marsh JL, Nepola JV, Dirschl DR, Hurwitz S, DeCoster TA.** Radiographic fracture assessments: which ones can we reliably make? J Orthop Trauma. 2000; 14(6): 379-85.
42. **Fleiss JL.** Statistical Methods for Rates and Proportions. New York: Wiley-Interscience; 1981.
43. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. Biometrics. 1977; 33(1): 159-74.
44. **Svanholm H, Starklint H, Gundersen HJ, Fabricius J, Barlebo H, Olsen S.** Reproducibility of histomorphologic diagnoses with special reference to the kappa statistic. APMIS. 1989; 97(8): 689-98.
45. **Petrie A, Sabin C.** Assessing agreement. In *Medical statistics at a glance*, pp. 105-107, Oxford: Blackwell Publishing; 2005.
46. **Rayan F, Dodd M, Haddad FS.** European validation of the Vancouver classification of periprosthetic proximal femoral fractures. J Bone Joint Surg Br. 2008; 90(12): 1576-9.

47. **Beaule PE, Dorey FJ, Matta JM.** Letournel classification for acetabular fractures. Assessment of interobserver and intraobserver reliability. J Bone Joint Surg Am. 2003; 85-A(9): 1704-9.
48. **Ploegmakers JJ, Mader K, Pennig D, Verheyen CC.** Four distal radial fracture classification systems tested amongst a large panel of Dutch trauma surgeons. Injury. 2007; 38(11): 1268-72.
49. **Belloti JC, Tamaoki MJ, Franciozi CE, Santos JB, Balbachevsky D, Chap Chap E, Albertoni WM, Faloppa F.** Are distal radius fracture classifications reproducible? Intra and interobserver agreement. Sao Paulo Med J. 2008; 126(3): 180-5.
50. **Sheps DM, Kiefer KR, Boorman RS, Donaghy J, Lalani A, Walker R, Hildebrand KA.** The interobserver reliability of classification systems for radial head fractures: the Hotchkiss modification of the Mason classification and the AO classification systems. Can J Surg. 2009; 52(4): 277-282.
51. **Morgan SJ, Groshen SL, Itamura JM, Shankwiler J, Brien WW, Kuschner SH.** Reliability evaluation of classifying radial head fractures by the system of Mason. Bull Hosp Jt Dis. 1997; 56(2): 95-8.
52. **Fung W, Jonsson A, Buhren V, Bhandari M.** Classifying intertrochanteric fractures of the proximal femur: does experience matter? Med Princ Pract. 2007; 16(3): 198-202.
53. **Jin WJ, Dai LY, Cui YM, Zhou Q, Jiang LS, Lu H.** Reliability of classification systems for intertrochanteric fractures of the proximal femur in experienced orthopaedic surgeons. Injury. 2005; 36(7): 858-61.
54. **van Embden D, Rhemrev SJ, Meylaerts SA, Roukema GR.** The comparison of two classifications for trochanteric femur fractures: the AO/ASIF classification and the Jensen classification. Injury. 2010; 41(4): 377-81.
55. **Hoesel LM, Pausch M, Schnettler R, Heiss C.** The impact of osteoporosis on the classification of hip and wrist fractures. Med Sci Monit. 2008; 14(3): HY1-8.
56. **Warriner AH, Patkar NM, Curtis JR, Delzell E, Gary L, Kilgore M, Saag K.** Which fractures are most attributable to osteoporosis? J Clin Epidemiol. 2011; 64(1): 46-53.
57. **Hayer PS, Deane AK, Agrawal A, Maheshwari R, Juyal A.** A Study on the Correlation of Pertrochanteric Osteoporotic Fracture Severity with the Severity of Osteoporosis. J Clin Diagn Res. 2016; 10(4): RC09-11.
58. **Dhainaut A, Daibes K, Odinsson A, Hoff M, Syversen U, Haugeberg G.** Exploring the relationship between bone density and severity of distal radius fragility fracture in women. J Orthop Surg Res. 2014; 9: 57.
59. **Beimers L, Kreder HJ, Berry GK, Stephen DJ, Schemitsch EH, McKee MD, Jaglal S.** Subcapital hip fractures: the Garden classification should be replaced, not collapsed. Can J Surg. 2002; 45(6): 411-4.
60. **Brunner A, Honigmann P, Treumann T, Babst R.** The impact of stereo-visualisation of three-dimensional CT datasets on the inter- and intraobserver reliability of the AO/OTA and Neer classifications in the assessment of fractures of the proximal humerus. J Bone Joint Surg Br. 2009; 91(6): 766-71.
61. **Martin JS, Marsh JL, Bonar SK, DeCoster TA, Found EM, Brandser EA.** Assessment of the AO/ASIF fracture classification for the distal tibia. J Orthop Trauma. 1997; 11(7): 477-83.
62. **Johnstone DJ, Radford WJ, Parnell EJ.** Interobserver variation using the AO/ASIF classification of long bone fractures. Injury. 1993; 24(3): 163-5.
63. **Dirschl DR, Adams GL.** A critical assessment of factors influencing reliability in the classification of fractures, using fractures of the tibial plafond as a model. J Orthop Trauma. 1997; 11(7): 471-6.
64. **Bernstein J, Adler LM, Blank JE, Dalsey RM, Williams GR, Iannotti JP.** Evaluation of the Neer system of classification of proximal humeral fractures with computerized tomographic scans and plain radiographs. J Bone Joint Surg Am. 1996; 78(9): 1371-5.
65. **Humphrey CA, Dirschl DR, Ellis TJ.** Interobserver reliability of a CT-based fracture classification system. J Orthop Trauma. 2005; 19(9): 616-22.

66. **Sjoden GO, Movin T, Guntner P, Aspelin P, Ahrengart L, Ersmark H, Sperber A.** Poor reproducibility of classification of proximal humeral fractures. Additional CT of minor value. Acta Orthop Scand. 1997; 68(3): 239-42.
67. **Harness NG, Ring D, Zurakowski D, Harris GJ, Jupiter JB.** The influence of three-dimensional computed tomography reconstructions on the characterization and treatment of distal radial fractures. J Bone Joint Surg Am. 2006; 88(6): 1315-23.
68. **Borrelli J, Jr., Goldfarb C, Catalano L, Evanoff BA.** Assessment of articular fragment displacement in acetabular fractures: a comparison of computerized tomography and plain radiographs. J Orthop Trauma. 2002; 16(7): 449-56; discussion 456-7.
69. **Cole RJ, Bindra RR, Evanoff BA, Gilula LA, Yamaguchi K, Gelberman RH.** Radiographic evaluation of osseous displacement following intra-articular fractures of the distal radius: reliability of plain radiography versus computed tomography. J Hand Surg Am. 1997; 22(5): 792-800.
70. **Katz MA, Beredjiklian PK, Bozentka DJ, Steinberg DR.** Computed tomography scanning of intra-articular distal radius fractures: does it influence treatment? J Hand Surg Am. 2001; 26(3): 415-21.
71. **Hu YL, Ye FG, Ji AY, Qiao GX, Liu HF.** Three-dimensional computed tomography imaging increases the reliability of classification systems for tibial plateau fractures. Injury. 2009; 40(12): 1282-5.
72. **Doornberg J, Lindenhovius A, Kloen P, van Dijk CN, Zurakowski D, Ring D.** Two and three-dimensional computed tomography for the classification and management of distal humeral fractures. Evaluation of reliability and diagnostic accuracy. J Bone Joint Surg Am. 2006; 88(8): 1795-801.
73. **Sjoden GO, Movin T, Aspelin P, Guntner P, Shalabi A.** 3D-radiographic analysis does not improve the Neer and AO classifications of proximal humeral fractures. Acta Orthop Scand. 1999; 70(4): 325-8.
74. **Sanders RW.** The Problem with Apples and Oranges. J Orthop Trauma 1997(7): 465-466.
75. **Swiontkowski MF, Sands AK, Agel J, Diab M, Schwappach JR, Kreder HJ.** Interobserver variation in the AO/OTA fracture classification system for pilon fractures: is there a problem? J Orthop Trauma. 1997; 11(7): 467-70.
76. **Craig WL, 3rd, Dirschl DR.** Effects of binary decision making on the classification of fractures of the ankle. J Orthop Trauma. 1998; 12(4): 280-3.
77. **Bernstein J, Monaghan BA, Silber JS, DeLong WG.** Taxonomy and treatment--a classification of fracture classifications. J Bone Joint Surg Br. 1997; 79(5): 706-7; discussion 708-9.
78. **Colton CL.** Fracture classification: A response to Bernstein et al. J Bone Joint Surg Br 1997(79): 708-709.
79. **Audige L, Bhandari M, Kellam J.** How reliable are reliability studies of fracture classifications? A systematic review of their methodologies. Acta Orthop Scand. 2004; 75(2): 184-94.
80. **Smith RM.** The classification of fractures. J Bone Joint Surg Br. 2000; 82(5): 625-6.
81. **Miller T, Kaeding CC, Flanigan D.** The classification systems of stress fractures: a systematic review. Phys Sportsmed. 2011; 39(1): 93-100.
82. **Bernstein J, Long JS, Veillette C, Ahn J.** Crowd intelligence for the classification of fractures and beyond. PLoS One. 2011; 6(11): e27620.
83. **Schneidmuller D, Roder C, Kraus R, Marzi I, Kaiser M, Dietrich D, von Laer L.** Development and validation of a paediatric long-bone fracture classification. A prospective multicentre study in 13 European paediatric trauma centres. BMC Musculoskelet Disord. 2011; 12: 89.