Radiographic staging of Kienböck's disease

Poor reproducibility of Ståhl's and Lichtman's staging systems

Claus Hjorth Jensen, Karsten Thomsen and Flemming Holst-Nielsen

We assessed the precision of 2 widely used radiographic classification systems for patients with Kienböck's disease by repeated staging of 76 radiographic examinations by 3 independent observers.

The intraobserver reproducibility was strong but the interobserver reliability was fair or moderate.

Hand Unit, Department of Orthopedics, Rigshospitalet, University of Copenhagen, DK-2100 Copenhagen, Denmark.
Correspondence: Dr. Claus Hjorth Jensen, Department of Orthopedics, Hvidovre Hospital, University of Copenhagen, DK-2650 Hvidovre, Denmark. Tel +45 36-363632. Fax -323782
Submitted 95-05-18. Accepted 96-01-29

Treatment of Kienböck's disease is linked to radiographic staging of the disease, thereby making the reproducibility of staging crucial (von Buck-Gramcko and Landkers 1990, Hooper 1992, Voche et al. 1992).

Three staging systems have been presented (Ståhl 1947, Decoulx et al. 1957, Lichtman et al. 1977), but no attempts have been made to assess their reliability. The precision of staging can be assessed by the observer variation, i.e., the interobserver reliability and the intraobserver reproducibility, and analyzed by kappa statistics (Cohen 1960, Gjørup and Jensen 1986).

We assessed the precision of the systems proposed by Ståhl and Lichtman et al., as these systems have been used most often.

Patients and methods

Radiographs obtained at referral and at follow-up of 48 patients with Kienböck's disease treated without surgery were studied. Each examination included an anteroposterior and a lateral projection. Rigid standardization, as recommended by Palmer (1987), was not used in obtaining the radiographs.

In 39 patients 2 examinations were available, in 8 patients only the radiographs obtained at the follow-up were available, and in 1 patient only the radiograph obtained at referral was available, as this patient refused to undergo further examinations. Thus 87 radiographic examinations were studied.

Staging according to Ståhl (1947), Lichtman et al. (1977) and Voche et al. (1992) was performed independently by 3 observers; 1 specialist in hand surgery and 2 experienced orthopedic surgeons, of whom 1 had a particular interest in hand surgery. Staging followed the guidelines given in the Ståhl and the Lichtman systems, respectively. Each observer was familiar with the staging systems. However, in order to standardize the information available to the observers, each was provided with a typed summary of the classification systems.

All patient-identifying data were obscured on the radiographs and after numbering the radiographs belonging together, the staging was performed in random order by each observer.

If one or more of the observers questioned the diagnosis of Kienböck's disease, the radiographs were excluded from the analysis.

The intraobserver reproducibility of the Ståhl staging system was estimated by repeated staging by the 3 observers after an interval of a minimum of 6 months.

Statistics

Kappa statistics, as described by Cohen (1960), allow for measuring observer agreement in dichotomous and rank scales taking into account the agreement by chance. Weighted kappa brings the magnitude of the disagreement between the individual assessments into the calculation. Weighted kappa can be calculated with different weights in accordance with the agreement expressed between stages (Gjørup and Jensen 1986). The present weighting was decided by the authors.

Perfect agreement between 2 runs was given the weight of 1. Disagreement concerning 1 stage between 2 runs was given the weight of 3/4. Disagreement concerning 2 stages between 2 runs was given the weight of 1/2. Disagreement concerning 3 stages between 2 runs was given the weight of 1/4, and disagreement concerning 4 stages between 2 runs was given the weight of 0.

Table 1. Strength of agreement given the value of kappa according to Landis and Koch (1977)

Value of kappa	Strength of agreement
< 0	poor
0–0.20	slight
0.21–0.40	fair
0.41–0.60	moderate
0.61–0.80	substantial
0.81–1.00	almost perfect

Table 2. Interobserver reliability expressed as kappa values after classification of the 76 sets of radiographs, according to Ståhl and Lichtman

Observer	Ståhl	Lichtman
Hand surgeon—orthopedic surgeon with interest in hand surgery	0.49	0.52
Hand surgeon—orthopedic surgeon	0.34	0.45
Orthopedic surgeon with interest in hand surgery—orthopedic surgeon	0.40	0.52

Table 3. Intraobserver reproducibility expressed as kappa values after 2 reviews of the 76 sets of radiographs classified according to Ståhl

Observer	
Hand surgeon	0.63
Orthopedic surgeon with interest in hand surgery	0.60
Orthopedic surgeon	0.26

Given p(o) = the observed agreement, and p(c) = the calculated expected random agreement, kappa was calculated:

Kappa = (p(o) – p(c)) / (1 – p(c)).

Landis and Koch (1977) have suggested guidelines in deciding the strength of agreement, given the value of kappa (Table 1).

Results

The initial 87 sets of radiographs were reduced, since 3 observers questioned the diagnosis of Kienböck's disease in 3 sets, 2 observers questioned the diagnosis in 4 sets, and 1 observer questioned the diagnosis in 4 sets. Thus, the analysis comprised 76 radiographic examinations.

The kappa value between the hand surgeon and the orthopedic surgeon with special interest in hand surgery corresponds to moderate strength of agreement in both the Ståhl and the Lichtman systems. The interobserver reliability between the hand surgeon and the orthopedic surgeon and the 2 orthopedic surgeons was also moderate in the Lichtman system. However, in the Ståhl system the agreement was only fair (Table 2).

The strength of agreement between the 2 runs was substantial for the hand surgeon, moderate for the orthopedic surgeon with a special interest in hand surgery and fair for the orthopedic surgeon (Table 3).

Discussion

Several classification systems have been subjected to critical analysis recently and have not proved valid. Thus the Lauge-Hansen staging system for ankle fractures (Thomsen et al. 1991), the Neer classification of proximal humeral fractures (Sidor et al. 1993) and the Garden classification of fractures of the femoral neck (Frandsen et al. 1988) have shown a disappointingly low reproducibility.

In our study on staging systems of Kienböck's disease, the kappa values of the interobserver reliability and the intraobserver reproducibility demonstrated that these staging systems are inadequate. We tried to imitate the everyday clinical situation by leaving the classification to orthopedic surgeons instead of including a radiologist. The differences in staging were clinically significant and would lead to different treatments.

The observed disagreements present an obstacle to the communication between researchers; epidemiological studies on patients with Kienböck's disease include staging of their disease.

A simpler classification system may offer a solution to the problem, but should be introduced only after its intraobserver repeatability and interobserver reliability are shown to be acceptable.

References

von Buck-Gramcko D, Lankers J. Ergebnisse in der Therapie der Mondbeinnekrose. Untersuchungen an 91 Patienten. Handchir Mikrochir Plast Chir 1990; 22: 28-38.

Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Measurement 1960; XX (1): 37-46.

Decoulx P, Marchand M, Minet P, Razemon J-P. La maladie de Kienböck chez le mineur. Lille-Chirurgical 1957; 12: 65-81.

Frandsen P A, Andersen E, Madsen F, Skjødt T. Garden's classification of femoral neck fractures. An assessment of interobserver variation. J Bone Joint Surg (Br) 1988; 70: 588-90.

Gjørup T, Jensen A M. Kappakoefficienten—et mål for reproducerbarhed af nominale og ordiale data. Nord Med 1986; 101 (3): 90-4.

Hooper G. Kienböck's disease, Editorial. J Hand Surg 1992; 17B: 3-4.

Landis J R, Koch G G. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159-74.

Lichtman D M, Mack G R, MacDonald R I, Gunther S F, Wilson J N. Kienböck's disease: The role of silicone replacement arthroplasty. J Bone Joint Surg (Am) 1977; 5: 899-908.

Palmer A K. Editorial: Kienböck's disease—The influence of arthrosis on ulnar variance determination. J Hand Surg 1987; 12B: 291-2.

Sidor M L, Zuckerman J D, Lyon T, Koval K, Cuomo F, Schoenberg N. The Neer classification system for proximal humeral fractures. J Bone Joint Surg (Am) 1993; 75: 1745-50.

Ståhl F. On lunatomalacia. Acta Chir Scand (Suppl 126) 1947; 1-133.

Thomsen N O B, Overgaard S, Olsen L H, Hansen H, Nielsen S T. Observer variation in the radiographic classification of ankle fractures. J Bone Joint Surg (Br) 1991; 73: 676-8.

Voche P, Bour C, Merle M. Scapho-trapezio-trapezoid arthrodesis in the treatment of Kienböck's disease—a study of 16 cases. J Hand Surg 1992; 17B: 5-11.