ORIGINAL PAPER
PŮVODNÍ PRÁCE

Comparative Assessment of Outcome in Osteoarthritis of the Knee:
The Utility of Knee Scores

Srovnání výsledků léčení u osteoartritidy kolena – použitelnost bodových systémů

S. KESSLER[1], W. KÄFER[2]

[1] Orthopaedic Department, District Hospital Sindelfingen-Böblingen, Sindelfingen, Germany
[2] Orthopaedic Department Ulm University, Ulm, Germany

ABSTRACT

PURPOSE OF THE STUDY

The utility of scoring systems, which are used to determine health status or treatment benefit in patients with knee osteoarthritis is under discussion. Therefore it was the purpose of our investigation to evaluate the reliability and the concordance of two established knee scoring systems.

METHODS

Thirty-eight patients with unilateral knee osteoarthritis were scored by the Hospital for Special Surgery score and the Knee Society score. Two blinded observers rated the patients independently in order to determine the concordance of the scores, the correlation between the overall scores and their subscales such as "pain", "function" and "range of motion" and the inter-observer and intra-observer reliability.

RESULTS

There was a high correlation between the overall scores (r= 0.80) and between the scores and their subscales "range of motion" (r=0.89) and "function" (r=0.74). The correlation of scores for "pain" was slightly less (r=0.61). Mean inter-observer reliability ranged between r=0.58 and r=0.61. Mean intra-observer reliability was high for the overall scores as well as for the subscales of both scoring sytems (r=0.64 to r=0.93 and r=0.73 to r=0.92).

CONCLUSION

We have found that the assessment of overall scores as well as of their main subscales is concordant and reliable in our patient sample. The application of these scoring systems in measuring health status in patients with knee osteoarthritis appears to be an acceptable method of audit. However, we feel that presentation of the results of knee scoring systems should include detailed information on the main subscales, since this allows for a better understanding of results.

Key words: osteoarthritis, knee score, utility, reliability.

INTRODUCTION

Scoring systems to measure health status before and after treatment are established tools for auditing medical procedures (5, 6, 10). Rheumatologists started to assess health in their patients by scoring systems nearly fifty years ago (12). Over the following decades, more detailed instruments were developed measuring health in multiple dimensions (8, 9). Many of these instruments have been shown to be highly valid and reliable.

Within the last decades orthopaedic surgeons have detected scoring systems as methods to assess the benefit of their surgical interventions. The reasons for this developement might be decreasing financial resources in a field of rising competition between surgeons and an exaggerated need of documentation of the surgical outcome to the public.

A variety of scoring systems to measure outcome in orthopaedic surgery is available. They were mainly established to audit results of total hip and knee arthroplasty. Several investigators however have questioned the benefit of these scoring systems in assessing orthopaedic patients, due to an insufficient reliability and to the inability to give detailed information about the parameter, which has caused the change (1, 2, 3, 4).

It was the purpose of our investigation to determine the concordance and reliability of the Hospital for Special Surgery (HSS) score (6, 10) and the Knee Society score (5), two well established knee scoring systems commonly used to assess results of knee arthroplasty.

PATIENTS AND METHODS

We recruited our patients within the preoperative preparations for a primary total knee replacement. All had symptomatic osteoarthritis of one knee joint and radiological changes, graded 3 to 4 according to Kellgren & Lawrence (7). Patients with severe general medical pro-

Figure 1. Median, 5th, 25th, 75th and 95th percentile of the HSS score at different times (r1 t1 = rater 1/time 1; r1 t2 = rater 1/time 2; r2 t1 = rater 2/time 1)

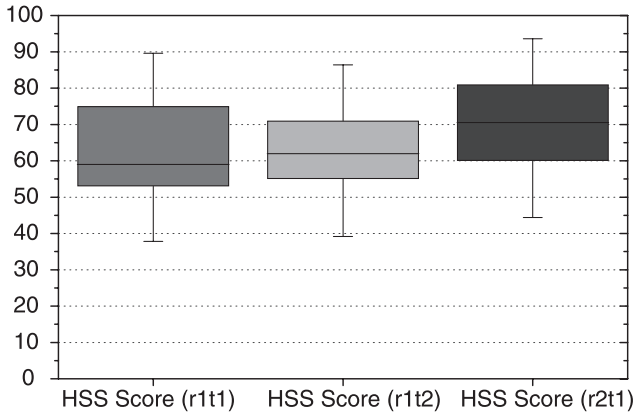


Figure 2. Median, 5th, 25th, 75th and 95th percentile of the Knee Society score at different times (r1 t1 = rater 1/time 1; r1 t2 = rater 1/time 2; r2 t1 = rater 2/time 1)

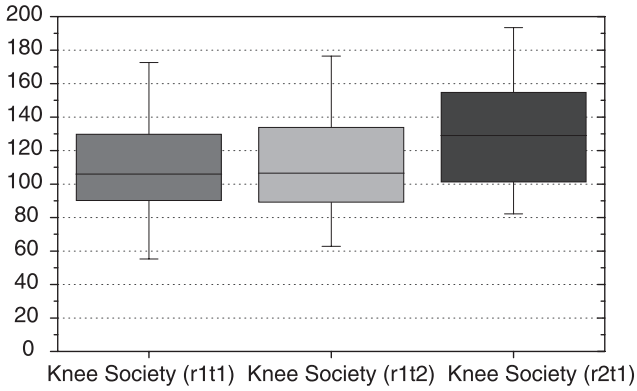
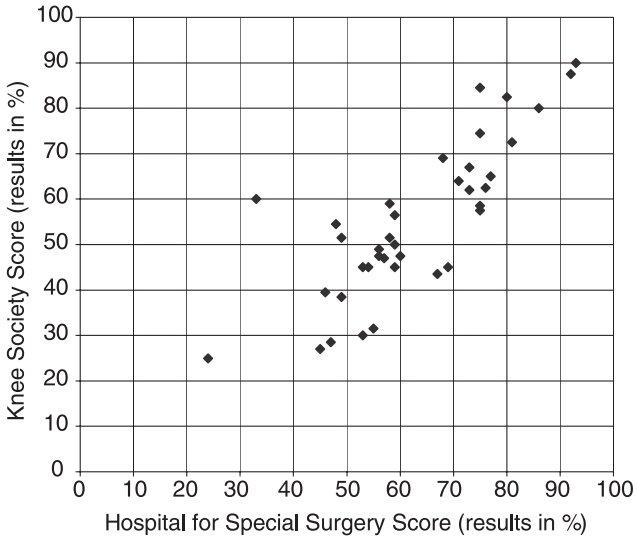


Figure 3. The correlation of the overall values of HSS Score and Knee Society score (n = 38)



blems, that could influence score results, were excluded.

All patients were assessed with a composite of the HSS score and the Knee Society score in an arbitrary manner, to avoid the possibility of an observer bias towards a particular system.

To determine the inter-observer reliability of both rating systems the investigations were independently performed by two different observers in the outpatient clinic on the same day. Intra-observer reliability was assessed by repeated assessment when the patients were admitted to the wards.

The HSS score measures "pain", "function" ("climbing stairs", "walking distance" and "transfer activity"), "range of motion", "muscle strength", "deformity", and "instability" (6, 10). It scores form a total of 100 points, representing an optimal health status. The subscales incuded have a different percentage impact to the overall score when compared to the Knee Society score.

The Knee Society score is scored out of 200 points, which again represents good health status. In this knee rating system the subscales "pain", "function" ("climbing stairs" and "walking distance") "range of motion" and "stability" are graded (5). To reduce the problem of a declining knee score result due to general medical problems, the Knee Society score is subdivided into a knee score that rates only the knee joint itself and a functional score that rates the patientįs ability to walk and climb stairs.

Data analysis was performed using the SAS package of statistical programs (11). To quantify the strength of concordance between both knee rating systems and to determine their inter- and intra-observer reliability the Spearman rank correlation coefficient was calculated for the overall scores and for the parameters "pain", "function" and "range of motion" seperately.

The relationship between the overall score of both rating systems was illustrated by a scatterplot. Median, 25th and 75th percentile as well as 5th and 95th percentile of the overall scores at different times were shown graphically in box & whisker diagrams.

RESULTS

Of 60 patients screened for primary total knee replacement for this investigation, 16 failed to meet the entry criteria due to general medical problems. Another six patients refused to participate. 38 patients finally took part in the study. 22 of them were female, 16 male. The mean age was 71 years, with a range from 59 to 81 years.

Correlation of rating systems

Results regarding median, 5th, 25th, 75th and 95th percentile of the overall scores of the HSS score and the Knee Society rating system are illustrated in figure 1 and 2 for different times.

Spearman rank correlation coefficient showed a good correlation for the overall scores (r=0.80) as well as for the parameters "range of motion" (r=0.89) and "function" (r=0.71). The correlation for "pain" was slightly less (r=0.61). All results were highly significant ($p<0.001$). The scatterplot shown in figure 3 illustrates the relationship between the overall scores.

Intra- and inter-observer reliability

Results regarding reliability of rating systems are listed in table 1 and 2. Intra-observer reliability was high for the overall scores of both rating systems and for parameters compared separately. Inter-observer reliability was slightly less. All results were highly significant ($p<0.001$).

Table 1. Inter- and intra-observer reliability (Spearman rank correlation coefficient) of the overall scores and the main parameters of the HSS score (= p<0.001)*

	Overall score	Pain	Range of motion	Function
Inter-observer-reliability	0.66*	0.58*	0.59*	0.62*
Intra-observer-reliability	0.80*	0.93*	0.64*	0.86*

Table 2. Inter- and intra-observer reliability (Spearman rank correlation coefficient) of the overall scores and the main parameters of the Knee Society score (= p<0.001)*

	Overall score	Pain	Range of motion	Function
Inter-observer-reliability	0.59*	0.59*	0.58*	0.66*
Intra-observer-reliability	0.89*	0.85*	0.73*	0.92*

DISCUSSION

Assessing "health status" or "treatment outcome" by rating systems is an established method of audit in rheumatology. A plenitude of rating systems is available, which often highly correlate and therefore provide evidence of concordance, suggesting that such instruments or their subscales are interchangeable to some degree (8).

Orthopaedic surgeons also have detected such rating systems as method for assessing the outcome of their surgically treated patients. However, the usefulness of these scores strongly has been questioned.

It was the purpose of our investigation to compare the results of the two such rating systems in patients with osteoarthritis of the knee. We have chosen the HSS score and the Knee Society score, which are commonly used scores, persuing that purpose in the last ten years.

We compared the overall scores and those subscales, which are considered to be the essential ones in measuring patients with osteoarthritis of the knee. We found a high correlation of both overall scores and their relevant subscales. Merely the parameter "pain" correlated slightly less. However, it is questionable, in how far how such a subjective parameter can be measured adequately, since the amount of pain highly depends on the patienţs actual perception, which is influenced by a variety of different factors (13). Other investigators also described "pain" to be difficult and inaccurate to score (2).

In another step we determined the reliability of these rating systems, seperately for their overall scores and for their main subscales and found high intra-observer reliability in all four cases. When determining inter-observer reliability, the correlation was slightly less. However, the variability between raters is still acceptable low.

Reviewing the literature, which deals with the utility of rating systems, two major points of criticism constantly can be found.

1. ***"Rating systems are not reliable in measuring heath status or treatment outcome"*** *(1, 3, 4, 13).*

Andersson (1) compared nine different hip scores in 77 patients after hemiarthroplasty and concluded, that the different scores would produce different results. Bryant et al. (3) tested 13 methods of hip scoring in 47 patients with hip arthroplasty; Callaghan and co-workers (4) examined 100 patients after total hip replacement with five different hip rating systems. They concluded analogously that the outcome of total hip replacement would strongly depend on the score used to measure that outcome. Tillman et al. (13), who compared three rating systems in 40 patients with knee osteoarthritis, reported a great variability in score results and a significant degree of inter-observer error.

These findings are in contrast to our results. Knowing about the problems deriving from the engagement in rating systems, we created a detailed manual within the planing process for this investigation, as a tool to standardize the investigation procedure, and to minimize the variability. The manual gives a precise description about the way the questions should be posed and the examination conducted. We think that these attempts to minimize variation are responsible for our good results.

2. ***"The description of health by overall score results and the comparison of overall score results is not reasonable, due to the different variable weight given to the parameters measured and due to the fact that a change of the overall score canţt tell the parameter which is responsible for that change"*** (3, 4, 13).

Callaghan and co-workers (4) following-up patients with total hip replacements recommended that "emphasis should be placed on important individual parameters rather than on overall scores". Bryant et al. (3), when investigating hip scores, identified three so called "core" variables: "pain", "function", and "range of motion". They showed that an assessment of more than these variables would bring only little additional information. They also suggested that a "three factor" score should be used in assessing outcome of hip arthroplasty seperately for these variables. Tillman et al. (13) concluded in a similar way, when comparing three different rating systems; they recommended separate statements for "pain", "function" and "range of motion".

We partially agree with these authors concerning the handling of scores, even although the overall values of the HSS score and the Knee Society score did correlate high in our study. One reason is the different weight of the analysed subscales. For example, 50% of the Knee Society score are dedicated to the parameter "function",

whereas only 22% in the HSS score are. Furthermore we are concerned that reporting the overall score only might miss the single subscale, which is responsible for the change in the patient`s health status. For example, when "range of motion" has improved distinctly after total knee replacement whereas "pain" has just slightly increased, the overall score will indicate an improvement, without showing the slight deterioration regarding the pain scale. This seems to be the essential point of criticism, because it should be one of the purposes of any audit to analyse which parameters cause improvement or deterioration.

Further work is necessary to create an ideal rating scale for patients with osteoarthritis of the knee.

ZÁVĚR

V práci autoři diskutují užitečnost dvou bodových systémů pro hodnocení zdravotního stavu a přínosu léčby u pacientů s osteoartritidou kolena. Cílem studie bylo stanovit spolehlivost a srovnatelnost dvou systémů bodového hodnocení kolena, které byly použity u jednoho souboru pacientů. Protože byla prokázána omezená užitečnost celkového skóre, autoři se domnívají, že výsledky naměřených parametrů by měly být prezentovány samostatně.

Soubor 58 pacientů s jednostrannou osteoartritidou kolena byl hodnocen pomocí skóre Nemocnice pro speciální chirurgii (Hospital for Special Surgery Score) a „Knee Society" skóre. Ke zjištění korelace mezi celkovým skóre a vlastními parametry kloubu, tj. bolest, funkce a rozsah pohybu, a k hodnocení jejich spolehlivosti byla měření prováděna nezávisle dvěma odborníky.

Byla zjištěna vysoká korelace mezi celkovým skóre obou hodnotících systémů (r = 0.80) i mezi parametry „rozsah pohybu" (r = 0.89) a „funkce" (r = 0.74). Korelace u parametru „bolest" byla poněkud nižší (r = 0.61). Spolehlivost výsledků mezi hodnotícími odborníky se v průměru pohybovala od r = 0.58 do r = 0.61. Spolehlivost hodnocení každého z hodnotitelů byla vysoká v obou systémech hodnocení jak u celkového skóre (r = 0.64 – r = 0.93), tak u jednotlivých parametrů (0.73 – r = 0.92).

Autoři zjistili, že ve studovaném souboru pacientů bylo u obou systémů celkové skóre souhlasné, stejně jako hlavní parametry kolenního kloubu, a dostatečně spolehlivé. Použití těchto bodových systémů pro hodnocení zdravotního stavu pacientů s artrózou kolena a jejich srovnávání se jeví jako metoda vhodná pro audit. Přesto se autoři domnívají, že by měly výsledky bodových hodnotících systémů být prezentovány pouze na základě hlavních parametrů, protože tento přístup dovoluje analýzu toho parametru, který vyvolal změnu.

References

1. ANDERSSON, G.: Hip Assessment: A comparison of nine different methods. J. Bone Jt Surg., 54-B: 621–625, 1972.
2. BREWSTER, S., NEWMAN, J. H.: Can knee replacements be assessed by post? Health Trends, 23: 113–114, 1991.
3. BRYANT, M. J., KERNOHAN, W. G., NIXON, J. R., MOLLAN, R. A. B.: A statistical analysis of hip scores. J. Bone Jt Surg., 75-B: 705–709, 1993.
4. CALLAGHAN, J. J., DYSART, S. H., SAVORY, C. F., HOPKINSON, W. J.: Assessing the Results of Hip Replacement. J. Bone Jt Surg., 72-B: 1008–1009, 1990.
5. INSALL, J. N., DORR, L. D., SCOTT, R. D., SCOTT, N.: Rationale of the knee society clinical rating system. Clin. Orthop., 248: 13–14, 1989.
6. INSALL, J. N., RANAWAT, C. S., AGLIETTI, P., SHINE, J.: A comparison of four models of total knee-replacement prostheses. J. Bone Jt Surg., 58-A: 754–765, 1976.
7. KELLGREN, J. H., LAWRENCE, J. S.: Radiological assessment of osteoarthritis. Ann. Rheum. Dis., 16: 494–502, 1957.
8. LIANG, H. L., LARSON, M. G., CULLEN, K. E., SCHWARTZ, J. A.: Comparative Efficiency and Sensitivity of five Health Status Instruments for Arthritis Research. Arthritis Rheum., 28: 542–547, 1985.
9. MEENAN, R. F., GERTMAN, P. M., MASON, J. H.: Measuring Health Status in Arthritis: The Arthritis Impact Measurement Scales. Arthritis Rheum., 23: 146–152, 1980.
10. RANAWAT, C. S., SHINE, J. J.: Duocondylar total knee arthroplasty. Clin. Orthop. 94: 185–195, 1973.
11. STATISTICAL ANALYSIS SYSTEM: SAS User's Guide. Cary, North Carolina, SAS Institute 1982.
12. STEINBROCKER, O., TRAEGER, C. H., BATTERMAN, R. C.: Therapeutic criteria in rheumatoid arthritis. J. Amer. Med. Ass. 140: 659–662, 1949.
13. TILLMAN, R. M., HARVEY, R. A., NOBLE, J., HIRST, P.: Total knee replacement – What's the score? An assessment of current methods of audit. Knee, 1: 65–68, 1994.

Priv.-Doz. Dr. med. Stefan Kessler,
Orthopaedic Department,
District Hospital Sindelfingen-Böblingen,
Arthur-Gruber Strasse 70,
710 65 Sindelfingen,
Germany
E-mail: stefan-kessler@t-online.de
Tel.: (+49) 7031 982481
Fax: (+49) 7031 982493

Práce byla přijata 23. 4. 2007.