CLINICAL ORTHOPAEDICS AND RELATED RESEARCH
Number 399, pp. 184–196

Scoring Systems in Total Knee Arthroplasty

Christian Michael Bach, MD; Michael Nogler, MD*; Iris Eva Steingruber, MD**; Michael Ogon, MD*; Cornelius Wimmer, MD*; Georg Göbel, MPh†; and Martin Krismer, MD**

For assessment of total knee arthroplasty outcome, various scoring systems have been introduced. The current study assessed the interobserver correlation of four commonly used total knee arthroplasty outcome scores. One hundred eighteen total knee arthroplasties were investigated by two independent observers, using the Hungerford score, the Hospital for Special Surgery score, the Knee Society score, and the Bristol score. Each score consisted of three subscores: pain, knee, and function. For the highest interobserver correlation was computed for the Bristol score (interobserver correlation coefficient, 0.88). For knee range of motion, flexion contracture, and extension lag there was high interobserver correlation (interobserver correlation coefficient > 0.8 each). For walking distance and walking aids, there also was a high interobserver correlation (interobserver correlation coefficient > 0.7 each). For clinical assessment of total knee arthroplasty, pain should be measured on a four-step system, the knee should be assessed by measurement of range of motion, extension lag, and flexion contracture, and function should be measured on a separate score assessing walking distance and walking aids.

From the *Departments of Orthopaedic Surgery, **Radiology, and †Biostatistics, University of Innsbruck, Innsbruck, Austria.

Reprint requests to Christian Michael Bach, MD, Department of Orthopaedic Surgery, University of Innsbruck, Anichstrasse 35, A-6020 Innsbruck, Austria.

Received: October 16, 2000.

Revised: May 14, 2001; September 23, 2001.

Accepted: October 4, 2001.

It is common to assess total arthroplasty outcome on the basis of scoring systems. The total outcome may be presented on a numerical point scale. Alternatively, a category system may be used. Scores with a category system frequently use the terms excellent, good, fair, or poor.[1,14,15,22,25,20,32,40,45,46,49] The lack of uniformity among analysis methods and the use of different terminology have been identified as problems.[3,4,7,8,12,26] For total hip replacement, outcome results were inconsistent, often giving contrary measures of success in the same patient.[7] Galante[12] suggested that a system for evaluating the results of hip surgery ideally should provide objective parameters that can be measured in a reproducible manner by independent observers.

The current study was designed to evaluate four commonly used scoring systems for knee arthroplasty: the Hungerford score, the Hospital for Special Surgery score, the Knee Society score, and the Bristol score. The study identifies variables proving high interobserver correlation which could be recommended for assessment of total knee arthroplasty.

MATERIALS A

A series of 118 tot
tients, 58 women an
their total knee arth
tion were included i
the patients at follo
years). The average
2–17 years).

All patients were
orthopaedic surgeon
monly used total kn
tems were used: the
for Special Surgery s
and the Bristol score
four scores for each
were recorded in a d
were started once d
The individual resu
completion of the pa
score outcome was c
dividual numerical
points (Bristol score
score). Statistical ana
mann test to compare
The individual scores

**TABLE 1. Inter
Total score Outc**

Subscore	
Pain	Pain
Knee	Mediol
	Antero
	Knee a
	Flexior
	Extens
	Flexion
	Extens
	Range
Function	Muscle
	Walkin
	Stairs
	Walkin
	Transfe
	Chair
	Giving
	Total

[1]Knee score of the Knee S
[2]Function score of the Kne

*MD**;
*fD**;
h†;

otion, extension lag, and nd function should be score assessing walking ids.

s total arthroplasty out-coring systems. The to-esented on a numerical ely, a category system with a category system as excellent, good, fair, [2,40,45,46,49] The lack of alysis methods and the ology have been identi-[3,12,26] For total hip re-sults were inconsistent, measures of success in lante[12] suggested that a he results of hip surgery e objective parameters n a reproducible manner vers.

vas designed to evaluate coring systems for knee gerford score, the Hospi-score, the Knee Society score. The study identi-high interobserver cor-be recommended for as-arthroplasty.

MATERIALS AND METHODS

A series of 118 total knee arthroplasties in 92 patients, 58 women and 34 men, who had followup of their total knee arthroplasty at the authors' institution were included in the study. The average age of the patients at followup was 70 years (range, 43–87 years). The average followup was 9.8 years (range, 2–17 years).

All patients were evaluated by two experienced orthopaedic surgeons (CMB and MN). Four commonly used total knee arthroplasty evaluation systems were used: the Hungerford score,[18] the Hospital for Special Surgery score,[35] the Knee Society score,[20] and the Bristol score.[28] Each observer completed all four scores for each patient independently. All data were recorded in a database. Statistical calculations were started once data acquisition was completed. The individual results were not discussed before completion of the patient examination forms. Total score outcome was computed in points using the individual numerical scale, which ranged from 50 points (Bristol score) to 200 points (Knee Society score). Statistical analysis included the global Friedmann test to compare the mean total point outcome. The individual scores were compared pairwise by the Wilcoxon test with post hoc Bonferroni correction. The results of the Bristol score were doubled and the function score and knee score of the Knee Society score were analyzed independently to achieve a maximum of 100 points for each score. Additionally, the results were categorized as excellent, good, fair, or poor outcome according to the individual score instructions. For each score, the total number of cases rated excellent or good by both observers and the total number of cases rated fair or poor by both observers were summarized. All four scores were divided additionally into three subscores: subscore pain, subscore knee, and subscore function.

The subscore pain consisted only of the variable pain in all four scores. The subscore knee contained variables such as mediolateral and anteroposterior stability, knee alignment, flexion contracture, extension lag, and range of motion (ROM). The subscore function included variables such as muscle strength, walking distance, ability to climb stairs, use of walking aids, transfer activity, chair, and giving way. The interobserver correlation was calculated for all variables using the nonparametric. Kendell Tau correlation coefficient (interobserver correlation coefficient) (Tables 1–5). All analyses were done using the SPSS statistical software

TABLE 1. Interobserver Correlation Coefficient for Subscore Variables and for Total score Outcome

		Interobserver Correlation Coefficient			
Subscore	**Variables**	*Hungerford Score*	*Hospital for Special Surgery Score*	*Knee Society Score*	*Bristol Score*
Pain	Pain	0.74	0.84	0.62	0.88
Knee	Mediolateral stability	0	0.38	0.07	
	Anteroposterior stability			0.53	
	Knee alignment	0.28	0.41	0.54	0.22
	Flexion contracture	0.93	0.93	0.93	
	Extension lag		0.85	0.87	
	Flexion contracture				0.84
	Extension lag				
	Range of motion	0.89	0.87	0.87	0.87
Function	Muscle strength	0.65	0.48		
	Walking distance		0.73	0.73	0.73
	Stairs		0.70	0.63	0.71
	Walking aids		0.79	0.81	0.81
	Transfer activity		0.42		
	Chair				0.76
	Giving way				0.54
	Total	0.65	0.82	0.48[1] 0.78[2]	0.84

[1]Knee score of the Knee Society score

[2]Function score of the Knee Society score

TABLE 2. Hungerford Score With Subscores

Score	Points	Interobserver Correlation Coefficient
Total score	100	0.65
Subscore pain	40	0.74
Frequent rest pain; Severe pain with any activity, necessitating nonweight-bearing status		
Analgesics stronger than aspirin required; Cannot do routine activities because of knee pain	0	
Common rest pain; Moderately severe pain with almost any activity; Heavy dependence on weightbearing assistance devices; Unable to work because of pain; Unable to do most work because of pain; Unable to do housework or even simple activities	10	
Less frequent rest pain; Routine activities moderately restricted; May require mild or narcotic analgesics	20	
Performs most routine activities with slight or tolerable pain; Slight or tolerable pain; Occasional mild, nonnarcotic analgesics or antiinflammatory medications	30	
No discomfort with routine activity; Excessive activity provokes discomfort; Rarely takes analgesic or antiinflammatory medication	35	
No pain with any activity	40	
Subscore Knee	50	
Mobility		0.89
1–30°	0	
31–60°	5	
61–90°	10	
91–105°	20	
>105°	25	
Flexion contracture		0.93
5–15°	−5	
15–30°	−10	
30–45°	−15	
>45°	−20	
Stability		0
0–5°	10	
5–15°	5	
>15°	0	
Varus or valgus deformity		0.28
0–5°	15	
6–10°	10	
11–15°	7	
16–20°	3	
>20°	0	
Subscore Function	10	
Quadriceps strength		0.65
>75%	10	
50–75%	5	
<50%	0	

TABLE 3. Hospita

Score
Total score
Subscore pain
No pain at anytime
No pain on walking
Mild pain on walking
Moderate pain on walking
Severe pain on walking
No pain at rest
Mild pain at rest
Moderate pain at rest
Severe pain at rest
Subscore Knee
Range of motion
1 point for each 8°
Instability
None
Mild: 0°–5°
Moderate: 6°–15°
Severe: 16° or more
Flexion deformity
None
A few degrees
5°–10°
11° or more
Alignment
Each 5° varus/valgus
Extension lag
5°
10°
15°
Crutches
One cane
One crutch
Two crutches
Subscore Function
Walking
Walking and standing unli
Walking distance of 5–10
Walking 1–5 blocks and si
Walking less than 1 block
Cannot walk
Stairs
Climbing stairs
Climbing stairs with suppo

nts	Interobserver Correlation Coefficient
0	0.65
)	0.74
)	
)	
)	
5	
)	
)	
	0.89
)	
)	
5	
	0.93
5	
0	
5	
0	
	0
)	
	0.28
5	
)	
)	
	0.65
)	

TABLE 3. Hospital for Special Surgery Score With Subscores

Score	Points	Interobserver Correlation Coefficient
Total score	100	0.82
Subscore pain	30	0.84
No pain at anytime	30	
No pain on walking	15	
Mild pain on walking	10	
Moderate pain on walking	5	
Severe pain on walking	0	
No pain at rest	15	
Mild pain at rest	10	
Moderate pain at rest	5	
Severe pain at rest	0	
Subscore Knee	38	
Range of motion		0.87
1 point for each 8°	≤ 18	
Instability		0.38
None	10	
Mild: 0°–5°	8	
Moderate: 6°–15°	5	
Severe: 16° or more	0	
Flexion deformity		0.93
None	10	
A few degrees	8	
5°–10°	5	
11° or more	0	
Alignment		0.41
Each 5° varus/valgus	−1	
Extension lag		0.85
5°	−2	
10°	−3	
15°	−5	
Crutches		0.79
One cane	−1	
One crutch	−2	
Two crutches	−3	
Subscore Function	32	
Walking		0.73
Walking and standing unlimited	12	
Walking distance of 5–10 blocks and standing ability intermittent (< 1/2 hour)	10	
Walking 1–5 blocks and standing ability up to 1/2 hour	8	
Walking less than 1 block	4	
Cannot walk	0	
Stairs		0.70
Climbing stairs	5	
Climbing stairs with support	2	

(continues)

TABLE 3. Hospital for Special Surgery Score With Subscores (*Continued*)

Score	Points	Interobserver Correlation Coefficient
Transfer		0.42
Transfer activity	5	
Transfer activity with support	2	
Muscle strength		0.48
Excellent		
Cannot break the quadriceps power	10	
Good		
Can break the quadriceps power	8	
Fair		
Moves through the arc of motion	4	
Poor		
Cannot move through the arc of motion	0	

TABLE 4. Knee Society Score Including the Knee Score and the Function Score

Score	Points	Interobserver Correlation Coefficient
Total score	200	
Subscore pain	50	0.62
None	50	
Mild or occasional	45	
Stairs only	20	
Walking and stairs	40	
Moderate	20	
Occasional	10	
Continual	0	
Severe		
Subscore Knee	50	0.48
Range of motion		0.87
One point for each 5°	≤25	
Anteroposterior stability		0.53
<5 mm	10	
5–10 mm	5	
>10 mm	0	
Mediolateral stability		0.07
<5°	15	
6°–9°	10	
10°–14°	5	
>15°	0	
Flexion contracture		0.93
5°–10°	−2	
10°–15°	−5	
16°–20°	−10	
>20°	−15	

(continues)

TABLE 4. Knee

Score
Extension lag
<10°
10°–20°
>20°
Alignment
5°–10°
0°–4°
11°–15°
>15°
Subscore Function
Walking
Unlimited
>10 blocks
5–10 blocks
< 5 blocks
Housebound
Unable
Stairs
Normal up and down
Normal up; down with r
Up and down with rail
Up with rail; down unat
Unable
Walking aids
Cane
Two canes
Crutches or walker

package (SPSS for Wir
Chicago, IL).

RESULTS

Total Score Outcom

The median total scor
for both observers an
74 points (range, 1
Hungerford score, 7:
points) for the Hospi
score, 58 points (ran
the knee score of the
points (range, −10–1
tion score of the Kne
points (range, 21–98
score (Fig 1). The gl

TABLE 4. Knee Society Score Including the Knee Score and the Function Score

Score	Points	Interobserver Correlation Coefficient
Extension lag		0.87
<10°	−5	
10°–20°	−10	
>20°	−15	
Alignment		0.54
5°–10°	0	
0°–4°	3/°	
11°–15°	3/°	
>15°	20	
Subscore Function	100	0.78
Walking		0.73
Unlimited	50	
>10 blocks	40	
5–10 blocks	30	
< 5 blocks	20	
Housebound	10	
Unable	0	
Stairs		0.63
Normal up and down	50	
Normal up; down with rail	40	
Up and down with rail	30	
Up with rail; down unable	15	
Unable	`0	
Walking aids		0.81
Cane	−5	
Two canes	−10	
Crutches or walker	−20	

package (SPSS for Windows, Version 8.0, SPSS, Chicago, IL).

RESULTS

Total Score Outcome

The median total score outcome, as averaged for both observers and for all the knees, was 74 points (range, 15–100 points) for the Hungerford score, 75 points (range, 23–97 points) for the Hospital for Special Surgery score, 58 points (range, −17–92 points) for the knee score of the Knee Society score, 50 points (range, −10–100 points) for the function score of the Knee Society score, and 71 points (range, 21–98 points) for the Bristol score (Fig 1). The global Friedmann test revealed significant differences among the mean total score outcomes ($p < 0.001$). Both components of the Knee Society rating system (knee score and function score) revealed significantly lower total score outcomes ($p < 0.05$, each).

Interobserver Correlation for the Total Point Outcome

The mean differences in points between Observer 1 and Observer 2 of the total score outcome were 14.1 points for the knee score of the Knee Society, 8.8 points for the Hungerford score, 7.2 points for the function score of the Knee Society score, 4.2 points for the Bristol score, and 3.9 points the Hospital for Special Surgery score.

TABLE 5. Bristol Score With Subscores

Score	Points	Interobserver Correlation Coefficient
Total score	50	0.84
Subscore pain	15	0.88
No pain	15	
Mild or occasional pain	12	
Moderate pain restricting activity	6	
Severe pain disturbing rest	0	
Subscore Knee	15	
Movement		0.87
1 point for each 12°	≤ 10	
Fixed flexion contracture or extension lag		0.84
None	3	
A few degrees	2	
Up to 15°	1	
16° or more	0	
Maximum varus/valgus		0.22
None	2	
up to 15°	1	
16° or more	0	
Subscore Function	20	
Mobility		0.81
No aids	5	
1 cane sometimes	4	
1 cane always	3	
2 canes	2	
2 crutches or frame	1	
Chairbound or bedridden	0	
Walking		0.73
Unlimited	5	
1/2 to 1 mile	4	
Up to 1/4 mile	3	
100 to 200 yards	2	
A few steps	1	
Cannot walk	0	
Stairs		0.71
Climbs stairs normally	3	
One at a time	2	
Only with help	1	
Cannot climb stairs	0	
Chair		0.76
Rises not using arms	2	
Rises using arms	1	
Cannot rise from chair	0	
Giving way		0.54
No giving way	5	
Feels insecure	3	
Gives way	2	
Does not support weight	0	

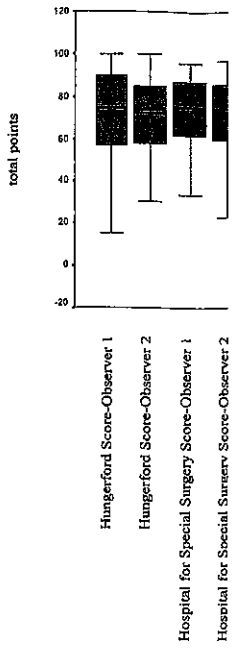


Fig 1. The total point
ford score, the Hosp
score, the Knee score
the Knee Society scor
Observer 1 and Obser
plot indicates the media
enty-fifth percentiles, a

The interobserver
was 0.84 for the Br
Hospital for Special
the function score of
0.65 for the Hungerfo
knee score of the Kn
1–5). Figure 2 show
server correlations fo

Total Score Categor

Evaluation of the sco
guish an excellent or
poor result showed ag
tween Observer 1 and
pital for Special Surg
disagreement was see
dred nine cases of a
for the Bristol score, 1
score of the Knee So

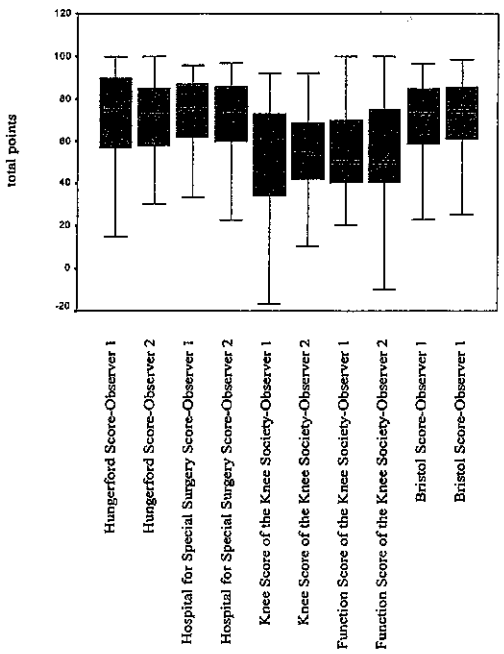


Fig 1. The total point outcomes for the Hungerford score, the Hospital for Special Surgery score, the Knee score and the Function score of the Knee Society score, and the Bristol score for Observer 1 and Observer 2 are shown. The boxplot indicates the median, the twenty-fifth, and seventy-fifth percentiles, and the adjacent values.

The interobserver correlation coefficient was 0.84 for the Bristol score, 0.82 for the Hospital for Special Surgery score, 0.78 for the function score of the Knee Society score, 0.65 for the Hungerford score, and 0.48 for the knee score of the Knee Society score (Tables 1–5). Figure 2 shows the different interobserver correlations for the scores tested.

Total Score Category Outcome

Evaluation of the scores' capability to distinguish an excellent or good result from a fair or poor result showed agreement in 112 cases between Observer 1 and Observer 2 for the Hospital for Special Surgery score. Accordingly, disagreement was seen in six cases. One hundred nine cases of agreement were recorded for the Bristol score, 103 cases for the function score of the Knee Society score, 96 cases for the Hungerford score, and 90 cases for the knee score of the Knee Society.

Interobserver Correlation of the Subscore Variables

Evaluation of interobserver correlation for the subscore variable pain revealed an interobserver correlation coefficient of 0.88 (Bristol score), 0.84 (Hospital for Special Surgery score), 0.74 (Hungerford score), and 0.62 (Knee Society score).

Variables of the subscore knee with a high interobserver correlation were ROM (interobserver correlation coefficient, 0.9 for all scores), flexion contraction (interobserver correlation coefficient, 0.9 for the Hungerford score, for the Hospital for Special Surgery score, and for the Knee Society score), extension lag (interobserver correlation coefficient, 0.9 for the Hospital for Special Surgery score and for the Knee Society score; interobserver correlation coefficient, 0.8 for the Bristol score).

For the subscore function, variables with high interobserver correlation were walking distance (interobserver correlation coefficient, 0.7 for the Hospital for Special Surgery score, for the Knee Society score, and for the Bristol score), the variable walking aids (interobserver correlation coefficient, 0.8 for the Hospital for Special Surgery score, for the Knee Society score, and for the Bristol score), chair (interobserver correlation coefficient, 0.8 for the Bristol score), stairs (interobserver correlation coefficient, 0.7 for the Hospital for Special Surgery score, for the Knee Society score, and for the Bristol score). The complete results are summarized in Tables 1 through 5.

DISCUSSION

Outcome assessment has received new impetus during the past decade as emphasis has shifted from the area of expansion and technical development to assessment and accountability.[2] An outcome scale for total joint arthroplasty should be valid, reliable, and responsive.[23] For total hip replacements, studies investigated various scoring systems and showed measurable dis-

orrelation Coefficient

0.84
0.88
0.87
0.84
0.22
0.81
0.73
0.71
0.76
0.54

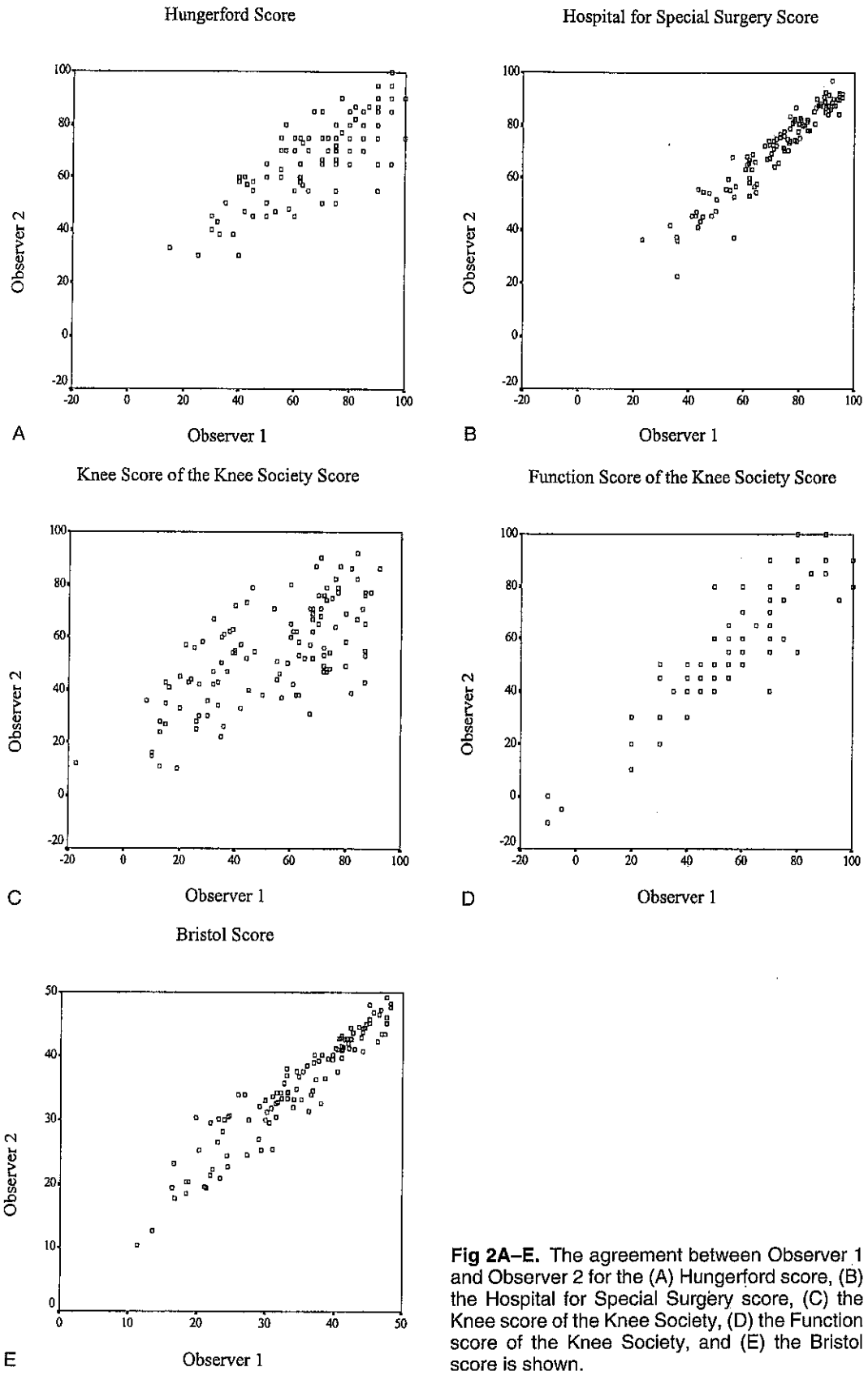


Fig 2A–E. The agreement between Observer 1 and Observer 2 for the (A) Hungerford score, (B) the Hospital for Special Surgery score, (C) the Knee score of the Knee Society, (D) the Function score of the Knee Society, and (E) the Bristol score is shown.

crepancies in the outc
arthroplasty, the respc
scores was investigated
et al[38] addressed the p
variability of knee scor
scores were unreliable
ternatives. The four kn
revealed considerable
come. The median tota
between 50 points for t
Knee Society and 75 pc
Special Surgery score.
that using different sc
difficult to compare res
reporting on outcome o
These results are in co
done on the evaluation
arthroplasty.[3,8]

For assessment of
outcome, it is commo
in categories like ex
poor.[1,14,15,22,25,30,32,40,4
these categories are e
summarizing excellen
fair and poor results.
[29,31–33,39] The current
considerable disagreen
servers when this simpl
pecially for the Hunger
score of the Knee So
that the practice of cat
developed as a matter o
as a result of the scienti
of the current study su
have a poor capability t
between excellent or g
poor results, respectiv
Special Surgery score a
vealed the highest inter
the category classificati

For total score outc
Special Surgery score
had the highest correla
plained by the subscor
vidual scores. Both sc
with high interobserv
1–5). Figure 2 shows
server correlation of th

crepancies in the outcome.[3,7,8] For total knee arthroplasty, the responsiveness of outcome scores was investigated by Kreibich et al.[23] Ryd et al[38] addressed the problem of interobserver variability of knee scores. They stated that the scores were unreliable but did not suggest alternatives. The four knee cores currently tested revealed considerable differences of total outcome. The median total score outcome ranged between 50 points for the function score of the Knee Society and 75 points for the Hospital for Special Surgery score. These results indicate that using different scoring systems makes it difficult to compare results of different studies reporting on outcome of total knee arthroplasty. These results are in concordance with studies done on the evaluation of scores for total hip arthroplasty.[3,8]

For assessment of total knee arthroplasty outcome, it is common to express the results in categories like excellent, good, fair, or poor.[1,14,15,22,25,30,32,40,45,46,49] In some studies these categories are even more simplified by summarizing excellent and good results and fair and poor results.[5,9,10,15,15,16,21,22,24,25,27,29,31–33,39] The current results show there was considerable disagreement between both observers when this simplification was applied especially for the Hungerford score and the knee score of the Knee Society. Gartland[13] stated that the practice of category classification has developed as a matter of convention rather than as a result of the scientific method. The results of the current study suggest that some scores have a poor capability to distinguish accurately between excellent or good results and fair or poor results, respectively. The Hospital for Special Surgery score and the Bristol score revealed the highest interobserver agreement for the category classification system.

For total score outcome, the Hospital for Special Surgery score and the Bristol score had the highest correlation. This might be explained by the subscore variables of the individual scores. Both scores include variables with high interobserver correlation (Tables 1–5). Figure 2 shows the different interobserver correlation of the five scores.

A weakness of the current study was having only two observers. For studies assessing the reliability of agreement of radiographic scoring systems multiple observer agreement was reported.[11,34,41–43,47,48] In clinical studies, repeated patient examinations by different observers using multiple evaluations systems may be problematic because of learning effects and decrease of patient cooperation.

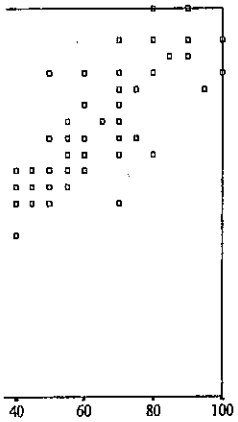
Pain is a critical issue that relates to the success or failure of a total knee replacement. Pain was the most important variable in all scores. Pain had the highest interobserver correlation in the Bristol score (interobserver correlation coefficient, 0.9) and in the Hospital for Special Surgery score (interobserver correlation coefficient, 0.8). In the Bristol score, pain is assessed by a simple four-point scale (no pain, mild or occasional pain, moderate pain restricting activity, and severe pain disturbing rest). The Hospital for Special Surgery score also uses a four-point scale. The other scores assess pain using a more complex system. The Hungerford score uses a six-point scale, the Knee Society score uses a seven-point scale where many different factors are considered for a proper ranking. The results of the current study suggest that for a high interobserver correlation, pain should be assessed using the four-point scales of the Bristol score or the Hospital for Special Surgery score. Similar suggestions were made by Aichroth et al[2] who introduced a standard chart for assessment of the knee after reconstructive operations. For assessment of pain, four simple grades with no pain at best, severe pain at worst, and two categories between were suggested. However, there are two additional approaches that could provide evaluation of pain: a visual analog scale as described by Huskisson[19] or measurement of the relief of pain instead of its severity. Neither of these methods currently is used for assessment of results of total knee arthroplasty.

For the subscore knee, the variables ROM, flexion contracture, and extension lag revealed the highest interobserver correlation (Tables 1–5). All these variables were measured in the

cial Surgery Score

ver 1

e Knee Society Score



rver 1

nt between Observer 1
) Hungerford score, (B)
Surgery score, (C) the
ociety, (D) the Function
ty, and (E) the Bristol

sagittal plane using a simple goniometer. The lateral malleolus of the fibula, the lateral femoral condyle, and the greater trochanter were clear landmarks for orientation and measurement. Similar results were reported for simple ROM measurements of the foot,[36] the wrist,[44] and for the knee.[38] Low interobserver correlations were seen for the variable knee alignment. A possible explanation may be the complicated measurement technique. A cord was required to measure the line from the center of the femoral head, to the center of the patellar, and to the center of the ankle. Additional use of a goniometer was difficult for one observer.

Determining the femoral head position clinically and assessing ROM using a goniometer may cause some inaccuracy. Ritter and Campbell[37] showed that the distance between the anterior iliac spines to be the only clinically significant factor related to the position of the femoral head center. Horton and Reckling[17] found that a marker placed just lateral to the palpable femoral pulse approximately 2 to 3 cm below the inguinal ligament is suitable as a guide to locate the center of the femoral head when determining the mechanical axis during total knee arthroplasty. Boone et al[6] showed that when more than one tester measures the same movement, increases in joint motion should exceed 6° for the lower extremity to determine improvement. With additional radiologic assessment accuracy probably would increase. However all scores applied in the current study assessed the knee alignment entirely clinically which may contribute to the low interobserver correlation.

The measurement of mediolateral knee stability revealed the lowest interobserver correlation of all variables tested (interobserver correlation coefficient values ranging from 0 to 0.4). As scores used steps of 5° for assessment of knee stability and a goniometer was required for proper measurement, it was difficult to test and to measure simultaneously.

The variables of the subscore function mainly revealed high interobserver correlation. The Knee Society score separates patients' general factors from knee related factors, whereas the other scoring systems combine both factors, and as such, may give a false impression of what is transpiring. As most of these variables were patient-dependent, a learning effect must be considered. Once estimated, the variables were easy to remember by the patients, which would result in a better interobserver correlation.

Uniform reporting would simplify interpretation of data and facilitate comparisons between results from different centers.[12] Orthopaedic surgeons should agree to a uniform method of evaluating and reporting the results of total knee arthroplasty. A system for evaluating the results of total knee replacement should provide objective parameters that can be measured in a reproducible manner by independent observers. Pain should be assessed on a four-step scale. The objective findings in the knee should include ROM, extension lag, and flexion contracture and evaluation of function should include walking distance and walking aides.

References

1. Aglietti P, Buzzi R, De Felice R, Giron F: The Insall-Burstein total knee replacement in osteoarthritis: A 10-year minimum follow-up. J Arthroplasty 14:560–565, 1999.
2. Aichroth PM, Freeman MAR, Smillie WA, Souter WA: A knee function assessment chart: From the British Orthopaedic Association Research Sub-Committee. J Bone Joint Surg 60B:308–309, 1978.
3. Andersson G: Hip assessment: A comparison of nine different methods. J Bone Joint Surg 54B:621–625, 1972.
4. Apley AG: An assessment of assessment [editorial]. J Bone Joint Surg 72B:957–958, 1990.
5. Baldwin JL, El Saied MR, Rubinstein Jr RA: Uncemented total knee arthroplasty: Report of 109 titanium knees with cancellous-structured porous coating. Orthopedics 19:123–130, 1996.
6. Boone DC, Azen SP, Lin CM, et al: Reliability of goniometric measurements. Phys Ther 58:1355–1390, 1978.
7. Bryant MJ, Kernohan WG, Nixon JR, Mollan RA: A statistical analysis of hip scores. J Bone Joint Surg 75B:705–709, 1993.
8. Callaghan JJ, Dysart SH, Savory CF, Hopkinson WJ: Assessing the results of hip replacement: A comparison of five different rating systems. J Bone Joint Surg 72B:1008–1009, 1990.
9. Cloutier JM, Sabouret P, Deghrar A: Total knee arthroplasty with retention of both cruciate ligaments: A nine to eleven-year follow-up study. J Bone Joint Surg 81A:697–702, 1999.
10. Diduch DR, Insall JN, Scott WN, et al: Total knee replacement in young, follow-up and functional 79A:575–582, 1997.
11. Flikkila T, Nikkola-Sihl interobserver reliability tures of the distal radiu mography is of minor 80B:670–672, 1998.
12. Galante J: The need fo evaluating results of tot Bone Joint Surg 67A:51
13. Gartland JJ: Orthopaedi cies in experimental de outcome. J Bone Joint S
14. Goldman RT, Scuderi C plantation for infected t Orthop 331:118–124, 19
15. Haas SB, Insall JN, M RE: Revision total kne modular components wi ment. J Bone Joint Surg
16. Hofmann AA, Kane KR of infected total knee ar ing spacer. Clin Orthop
17. Horton GA, Reckling F to the mechanical axis Arthroplasty 10:780–78
18. Hungerford DS, Kenn porous-coated anatomi North Am 13:103–122,
19. Huskisson EC: Meas 2:1127–1131, 1974.
20. Insall JN, Dorr LD, Sco of the Knee Society cli thop 248:13–14, 1989.
21. Karbowski A, Eckardt Revision resurfacing k loosening. Arch Orthop 1998.
22. Karthaus RP, Novakova in haemophilic arthro 70B:382–385, 1988.
23. Kreibich DN, Vaz M, E best way of assessing placement? Clin Orthop
24. Le AX, Cameron HU, C the patella following to pedics 22:395–398, 199
25. Li PL, Zamora J, Bentle of the Insall-Burstein II ical, radiological and s Joint Surg 81B:647–65
26. Liang MH, Katz JN, Pl arthroplasty outcome e can Academy of Orthop nominal group process. Orthopaedic Surgeons ies. J Bone Joint Surg 7
27. Lonner JH, Pedlow F arthroplasty for post-tr plasty 14:969–975, 199
28. Mackinnon J, Young sledge for unicompart

er scoring systems com-
as such, may give a false
s transpiring. As most of
atient-dependent, a learn-
isidered. Once estimated,
y to remember by the pa-
esult in a better interob-

would simplify interpre-
cilitate comparisons be-
ifferent centers.[12] Ortho-
ıld agree to a uniform
nd reporting the results of
. A system for evaluating
replacement should pro-
ters that can be measured
ıner by independent ob-
e assessed on a four-step
ıdings in the knee should
on lag, and flexion con-
n of function should in-
and walking aides.

Felice R, Giron F: The Insall-
ıcement in osteoarthritis: A 10-
ıp. J Arthroplasty 14:560–565,

n MAR, Smillie WA, Souter
assessment chart: From the
Association Research Sub-
nt Surg 60B:308–309, 1978.
ssment: A comparison of nine
one Joint Surg 54B:621–625,

ıent of assessment [editorial].
:957–958, 1990.
MR, Rubinstein Jr RA: Unce-
ıroplasty: Report of 109 tita-
illous-structured porous coat-
3–130, 1996.
in CM, et al: Reliability of go-
ts. Phys Ther 58:1355–1390,

VG, Nixon JR, Mollan RA: A
ıip scores. J Bone Joint Surg

I, Savory CF, Hopkinson WJ:
f hip replacement: A compar-
ating systems. J Bone Joint
1990.
P, Deghrar A: Total knee
ıtion of both cruciate liga-
-year follow-up study. J Bone
2, 1999.
Scott WN, et al: Total knee

replacement in young, active patients: Long-term follow-up and functional outcome. J Bone Joint Surg 79A:575–582, 1997.
11. Flikkila T, Nikkola-Sihto A, Kaarela O, et al: Poor interobserver reliability of AO classification of fractures of the distal radius: Additional computed tomography is of minor value. J Bone Joint Surg 80B:670–672, 1998.
12. Galante J: The need for a standardized system for evaluating results of total hip surgery [editorial]. J Bone Joint Surg 67A:511–512, 1985.
13. Gartland JJ: Orthopaedic clinical research: Deficiencies in experimental design and determinations of outcome. J Bone Joint Surg 70A:1357–1364, 1988.
14. Goldman RT, Scuderi GR, Insall JN: 2-stage reimplantation for infected total knee replacement. Clin Orthop 331:118–124, 1996.
15. Haas SB, Insall JN, Montgomery III W, Windsor RE: Revision total knee arthroplasty with use of modular components with stems inserted without cement. J Bone Joint Surg 77A:1700–1707, 1995.
16. Hofmann AA, Kane KR, Tkach TK, et al: Treatment of infected total knee arthroplasty using an articulating spacer. Clin Orthop 321:45–54, 1995.
17. Horton GA, Reckling FW: Femoral pulse as a guide to the mechanical axis in total knee arthroplasty. J Arthroplasty 10:780–784, 1995.
18. Hungerford DS, Kenna RV, Krackow KA: The porous-coated anatomic total knee. Orthop Clin North Am 13:103–122, 1982.
19. Huskisson EC: Measurement of pain. Lancet 2:1127–1131, 1974.
20. Insall JN, Dorr LD, Scott RD, Scott WN: Rationale of the Knee Society clinical rating system. Clin Orthop 248:13–14, 1989.
21. Karbowski A, Eckardt A, Schwitalle M, Heine J: Revision resurfacing knee arthroplasty for aseptic loosening. Arch Orthop Trauma Surg 117:256–258, 1998.
22. Karthaus RP, Novakova IR: Total knee replacement in haemophilic arthropathy. J Bone Joint Surg 70B:382–385, 1988.
23. Kreibich DN, Vaz M, Bourne RB, et al: What is the best way of assessing outcome after total knee replacement? Clin Orthop 331:221–225, 1996.
24. Le AX, Cameron HU, Otsuka NY, et al: Fracture of the patella following total knee arthroplasty. Orthopedics 22:395–398, 1999.
25. Li PL, Zamora J, Bentley G: The results at ten years of the Insall-Burstein II total knee replacement: Clinical, radiological and survivorship studies. J Bone Joint Surg 81B:647–653, 1999.
26. Liang MH, Katz JN, Phillips C, et al: The total hip arthroplasty outcome evaluation form of the American Academy of Orthopaedic Surgeons: Results of a nominal group process. The American Academy of Orthopaedic Surgeons Task Force on Outcome Studies. J Bone Joint Surg 73A:639–646, 1991.
27. Lonner JH, Pedlow FX, Siliski JM: Total knee arthroplasty for post-traumatic arthrosis. J Arthroplasty 14:969–975, 1999.
28. Mackinnon J, Young S, Baily RA: The St Georg sledge for unicompartmental replacement of the knee: A prospective study of 115 cases. J Bone Joint Surg 70B:217–223, 1988.
29. Martin SD, Haas SB, Insall JN: Primary total knee arthroplasty after patellectomy. J Bone Joint Surg 77A:1323–1330, 1995.
30. Mokris JG, Smith SW, Anderson SE: Primary total knee arthroplasty using the Genesis Total Knee Arthroplasty System: 3- to 6-year follow-up study of 105 knees. J Arthroplasty 12:91–98, 1997.
31. Mont MA, Mathur SK, Krackow KA, et al: Cementless total knee arthroplasty in obese patients: A comparison with a matched control group. J Arthroplasty 11:153–156, 1996.
32. Mont MA, Mayerson JA, Krackow KA, Hungerford DS: Total knee arthroplasty in patients receiving workers' compensation. J Bone Joint Surg 80A:1285–1290, 1998.
33. Mont MA, Yoon TR, Krackow KA, Hungerford DS: Eliminating patellofemoral complications in total knee arthroplasty: Clinical and radiographic results of 121 consecutive cases using the Duracon system. J Arthroplasty 14:446–455, 1999.
34. Nielsen JO, Dons-Jensen H, Sorensen HT: Lauge-Hansen classification of malleolar fractures: An assessment of the reproducibility in 118 cases. Acta Orthop Scand 61:385–387, 1990.
35. Ranawat CS, Shine JJ: Duo-condylar total knee arthroplasty. Clin Orthop 94:185–195, 1973.
36. Resch S, Ryd L, Stenström A, et al: Measuring hallux valgus: A comparison of conventional radiography and clinical parameters with regard to measurement accuracy. Foot Ankle 16:267–270, 1995.
37. Ritter MA, Campbell ED: A model for easy location of the center of the femoral head during total knee arthroplasty. J Arthroplasty 3 (Suppl):S59-S61, 1988.
38. Ryd L, Karrholm J, Ahlvin P: Knee scoring systems in gonarthrosis: Evaluation of interobserver variability and the envelop of bias: Score Assessment Group. Acta Orthop Scand 68:41–45, 1997.
39. Schai PA, Suh JT, Thornhill TS, Scott RD: Unicompartmental knee arthroplasty in middle-aged patients: A 2- to 6- year follow-up evaluation. J Arthroplasty 13:365–372, 1998.
40. Shih LY, Sim FH, Pritchard DJ, et al: Segmental total knee arthroplasty after distal femoral resection for tumor. Clin Orthop 292:269–281, 1993.
41. Sidor ML, Zuckerman JD, Lyon T, et al: The Neer classification system for proximal humeral fractures: An assessment of interobserver reliability and intraobserver reproducibility. J Bone Joint Surg 75A:1745–1750, 1993.
42. Siebenrock KA, Gerber C: The reproducibility of classification of fractures of the proximal end of the humerus. J Bone Joint Surg 75A:1751–1755, 1993.
43. Smith SW, Meyer RA, Connor PM, et al: Interobserver reliability and intraobserver reproducibility of the modified Ficat classification system of osteonecrosis of the femoral head. J Bone Joint Surg 78A:1702–1706, 1996.
44. Solgaard S, Carlsen A, Kramhoft M, Petersen VS: Reproducibility of goniometry of the wrist. Scand J Rehabil Med 18:5–7, 1986.
45. Tankersley WS, Hungerford DS: Total knee arthro-

plasty in the very aged. Clin Orthop 316:45–49, 1995.
46. Thomason HC, Wilson FC, Lachiewicz PF, Kelley SS: Knee arthroplasty in hemophilic arthropathy. Clin Orthop 360:169–173, 2000.
47. Thomsen NO, Overgaard S, Olsen LH, et al: Observer variation in the radiographic classification of ankle fractures. J Bone Joint Surg 73B:676–678, 1991.
48. Ward WT, Vogt M, Grudziak JS, et al: Severin classification system for evaluation of the results of operative treatment of congenital dislocation of the hip: A study of intraobserver and interobserver reliability. J Bone Joint Surg 79A:656–663, 1997.
49. Windsor RE, Insall JN, Urs WK, et al: Two-stage reimplantation for the salvage of total knee arthroplasty complicated by infection: Further follow-up and refinement of indications. J Bone Joint Surg 72A:272–278, 1990.

Sympto
Access

Taisuke
Kohei N

**Separated ossicles at th
olus, the condition kno
sometimes a cause of a
theories regarding the
One theory proposes th
sion fracture attributab
talofibular ligament, w
proposes that it is the r
sification center. Howe
reported that os subfibu
ture and few reports
have been described re
subfibulare that may h
cessory ossification rath
ture was treated by the**

Separated ossicles of t
condition known as os
found in 1% of the
subfibulare is sometim
in which case it is call
fibulare.[1] There are tw
origin of os subfibular
sicles are caused by an

From the Department of Or
University, Shimane, Japan
Reprint requests to Taisuk
Orthopaedics, Shimane Me
cho, Izumo, Shimane 693–8
Received: March 8, 2001.
Revised: September 10, 20
Accepted: October 11, 200