

Indexes of Severity for Osteoarthritis of the Hip and Knee Validation – Value in Comparison with Other Assessment Tests

M.G. LEQUESNE, C. MERY, M. SAMSON and P. GERARD

Département de Rhumatologie de l'Hôpital Léopold Bellan, Paris, France

The index of severity for hip disease (ISH) was established, validated and appraised as a new assessment test for the trial of new drugs as well as for long-term follow-up of patients, and to help with future indications for surgery. The ISH deals with pain, maximum walking distance, and some activities of daily living. Inter-observer reproducibility is good (mean deviation 0.55 points; $p < 0.05$). In a short-term, double-blind crossover trial, the ISH, judged according to its power to distinguish between the active drug period and the placebo period, appears as one of the best assessment tests. In the long term, total hip prosthesis is most often justified when the ISH score reaches 10–12 points.

The index of severity for knee disease (ISK) was validated and appraised by the same statistical methods. Its value in non-steroidal anti-inflammatory drug (NSAID) or analgesic trials is lower than the value of the ISH. However, its use is still justified for that purpose, and for long-term follow-up of osteoarthritis of the knee.

M.G. Lequesne, Département de Rhumatologie, Hôpital Léopold Bellan, 7 rue du Texel, 75014, Paris, France.

In an original paper in 1972, Andersson (1) showed a dramatic variation in the value of nine different methods of hip assessment: with the same series of patients, good results were obtained in 97.5% of patients using the most “optimistic” method, and in only 35% using the most pessimistic method! The patients’ own opinion midway between these two figures – 67% reported good results.

Having performed many drug trials dealing mainly with osteoarthritis (OA) of the extremities, we soon ascertained that most of the tests proposed to assess a new drug are of unknown value. However, such tests are very important, both to evaluate new non-steroidal anti-inflammatory drugs (NSAIDs) in the short term and to follow up the course of the disease in the long term, as well as for the appraisal of disease modifying drugs. Since 1980, we have undertaken studies to validate and evaluate the main clinical tests for OA of the hip and of the knee.

PATIENTS AND METHODS

Osteoarthritis of the hip

We established our index of severity for hip disease (ISH) in 1980 (2) (Table I) and it was validated in 1981 (3). The reproducibility was assessed in a double-blind study including 38 patients, interviewed separately by two investigators.

To appraise the value of the ISH among other tests, our method of validation is the reverse of the common therapeutic trial: in the latter, the value of the new drug is the unknown, the assessment tests being based on sound and well-known values. In our validation assay, the difference in benefit between placebo and a classical NSAID provided the known basis (being well established for many years) and the value of the assessment test was the unknown. This value was appraised by the power of each test to distinguish the active drug period from the placebo period in a double-blind, crossover, randomised, short-term trial.

In this study of the value of the ISH, we used indomethacin (75 mg) versus placebo, for one week each, in 55 patients suffering from OA of the hip. Each ISH question was standardised in its formulation. A second formulation was written in case of misunderstanding of the first one. The other assessment tests evaluated in this trial were pain level using the Huskisson visual analogue scale (4), patient’s and investigator’s overall opinion, walking time, and range of abduction and flexion.

Table I. Index of severity for osteoarthritis of the hip

	Points
<i>I – Pain or discomfort</i>	
A. During nocturnal bedrest	
• only on movement or in certain positions	1
• without movement	2
B. Duration of morning stiffness or pain after getting up	
• less than 15 min	1
• 15 min or more	2
C. Remaining standing for 30 min increases pain	1
D. Pain on walking	
• only after walking some distance	1
• early after starting	2
E. Pain or discomfort in sitting position for 2 h	1
<i>II – Maximum distance walked</i>	
• More than 1 km, but limited	1
• About 1 km (about 15 min)	2
• From 500 to 900 m (about 8–15 min)	3
• From 300 to 500 m	4
• From 100 to 300 m	5
• Less than 100 m	6
• With one walking stick or crutch	+1
• With two walking sticks or crutches	+2
<i>III – Activities of daily living</i>	
• Can you put on socks by bending forward?	0 to 2
• Can you pick up an object from the floor?	0 to 2
• Can you go up and down a standard flight of stairs?	0 to 2
• Can you get into and out of a car?	0 to 2
<i>Point score:</i> easily	0
with difficulty	1 (or 0.5 or 1.5)
impossible	2

The ISH is used by our team and by the orthopaedic surgeons in the hospital and the scores are regularly compared. From this we determined the length of time required for training in order to obtain a correct ISH interview, and the score beyond which hip replacement is necessary.

Osteoarthritis of the knee

The same methods were used for validation of OA of the knee. Twenty-seven patients with femoro-tibial OA entered the trial (femoro-patellar OA is too irregular in pain and severity and it is not a good “model” for trials). We used tiaprofenic acid (600 mg/day) versus placebo, for one week each. The index of severity for knee disease (ISK) (Table II) was compared with the following other assessment tests: pain level, patient’s and investigator’s overall opinion, time for going up and down a standard flight of stairs, duration of morning stiffness, limitation of flexion, and pain on flexion and extension.

RESULTS

Osteoarthritis of the hip

Inter-observer reproducibility. The mean deviation between the ISH scores found by two separate observers was 0.55 points. The test for a matched-pairs series did not show any significant difference ($t = 1.95$; $p < 0.05$).

Table II. Index of severity for osteoarthritis of the knee

	Points	
<i>I – Pain or discomfort</i>		
A. } B. } C. } D. }	As for the hip (see Table I)	
E. When getting up from sitting position without the help of arms		1
<i>II – Maximum distance walked</i>		
As for the hip (see Table I)		
<i>III – Activities of daily living</i>		
• Can you go up a standard flight of stairs?	0 to 2	
• Can you go down a standard flight of stairs?	0 to 2	
• Can you squat?	0 to 2	
• Can you walk on uneven ground?	0 to 2	
<i>Point score:</i> as for the hip (see Table I)		

Nevertheless, this good agreement is only reached if the observers have sufficient training: a minimum of 30 ISH interviews, with correction by a “senior” investigator, are necessary as a training period. After adequate training, an ISH score can be obtained within 3–4 min.

Value of the ISH – its place among other assessment tests. In our double-blind, crossover, short-term trial, the power of each test to distinguish between the active drug period and the placebo period was calculated and yielded the following results:

Investigator’s overall opinion	$p < 0.001$
Index of severity (ISH)	$p < 0.001$
Pain level (visual analogue scale)	$p < 0.01$
Patient’s overall opinion	$p < 0.01$
Walking time	$p < 0.05$
Abduction, flexion	Not significant

These results show that the subjective or semi-subjective criteria are best. However, limitation of mobility merits consideration if the aim of the study is to follow the patient in the long term, either in the natural disease course or during a long-term trial of 3–5 years with an anti-OA drug (if it exists . . .).

ISH helps to decide whether total hip prosthesis is necessary. According to our observations and those of the orthopaedic team over a period of seven years, hip prosthesis is justified when the ISH score reaches 10–12 points. More generally, the handicap can be described according to the ISH score, which reflects the clinical severity of OA of the hip as follows:

Points	Handicap
>14	Extremely severe
11, 12, 13	Very severe
8, 9, 10	Severe
5, 6, 7	Moderate
1–4	Mild

Such a measurement allows avoidance of arbitrary decisions in hip replacement, which may be too soon (often the case) or too late.

Osteoarthritis of the knee

Inter-observer reproducibility. Twenty-four patients interviewed separately by two blinded observers led to the following conclusions:

- The maximum difference between the two observers was 1.5 points.
- The mean deviation was 0.146.
- The test for a small (<30) matched-pairs series did not show any significant differences ($t = 0.167$; $p > 0.05$).
- In practice, with a risk of 5%, the mean difference between the two observers is between 0–0.5 points.

The inter-observer agreement is satisfactory.

Value of the ISK – its place among other assessment tests. According to the power of each test to distinguish between the active drug period and the placebo period, we have found:

Pain level	$p < 0.004$
Investigator's overall opinion	$p < 0.006$
Patient's overall opinion	$p < 0.014$
Index of severity (ISK)	$p < 0.025$
Time for going up and down a standard flight of stairs	$p < 0.05$
Duration of morning stiffness	} Not significant
Limitation of flexion	
Pain on flexion and extension	

DISCUSSION

There are many methods of assessment for OA of the hip and knee, but these were all established by surgeons in order to evaluate the results of a given surgical treatment. Most of them were not validated, and their results are very contradictory, as Andersson has shown in nine different methods concerning the hip (1). We have found a dozen methods for OA of the knee, all from orthopaedic surgeons (5–15). Furthermore, Jonsson (16) compared three different methods and found some discrepancy, but in an acceptable proportion. In most of these methods, pain parameters account for $30 \pm 5\%$ of the total score; in our ISH and ISK, pain accounts for one-third of the whole. The main differences between our medical indexes and the surgical indexes are as follows:

- (i) The ISH and ISK are easy and quick to use;
- (ii) They reflect the pain and functional status of the patient in his/her daily life, without reference to "objective" measurements;
- (iii) They have been validated and their reproducibility is satisfactory;
- (iv) They yield finer score differences than the surgical indexes and can be used for assessment of results in trials of new drugs as well as in daily practice to evaluate medical treatment. On average, when evaluation is started with either an ISH or ISK between 8 and 12 points, administration of an NSAID reduces the score by 2–3 points within a few days.

The ISH and ISK are also useful in following up diseases of the hip and knee in the long term, and to assess the value of disease modifying (anti-OA) drugs in controlled trials versus placebo for 3–5 years.

CONCLUSIONS

In conclusion, let us look at how we can use some of the assessment tests in daily practice, besides new drug trials.

For follow-up and the results of treatment, we personally use three tests at each visit: (i) visual analogue scale of pain; (ii) ISH or ISK; (iii) self-assessment by the patient, who has to choose one of the seven categories listed in Table III on reading the list.

For appraisal of the handicap and possible surgical indication, we use the ISH and ISK and the self-assessment shown in Table IV. The comparison between the degree of handicap scored from the ISH or ISK and the self-assessment of handicap is very interesting. Often, the patient estimates himself one grade above the handicap indicated by our indexes. This seems natural. If he places himself two grades or more above the handicap as defined by ISH or ISK, one could suspect either a psychological problem or a neurotic personality.

In summary, ISH and ISK seem to be good methods for checking more exactly the actual handicap of the patient, and the possible benefit from a given treatment.

Table III. *Self-assessment of result*

How do you feel in comparison with your last visit:	
● Much better?	● A little worse?
● Better?	● Worse?
● A little better?	● Much Worse?
● Same?	

Table IV. *Self-assessment of the handicap*

How do you rate your handicap:	
● Almost unbearable?	● Moderate?
● Very severe?	● Mild?
● Severe?	

REFERENCES

- Andersson G. Hip assessment: a comparison of nine different methods. *J Bone Joint Surg* 1972; 54B: 621–5.
- Lequesne M, Mery C. European guidelines for clinical trials of new antirheumatic drugs. *EULAR Bull* 1980; 9: 171–5.
- Lequesne M, Samson M. A functional index for hip diseases. Reproducibility. Value for discriminating drug's efficacy. 15th International Congress of Rheumatology. Expansion Scientifique française, Paris, 1981: 778–9.
- Huskisson E. Measurement of pain. *Lancet* 1974; 4: 1427.
- Aichroth P, Freeman MAR, Smilie IS, Souter WA. A knee function assessment chart. *J Bone Joint Surg* 1978; 60B: 308–9.
- Bargren JH, Freeman MAR, Swanson SAV, Todd RC. ICLH (Freeman/Swanson) arthroplasty in the treatment of arthritic knee. *Clin Orthop* 1976; 120: 65.
- Evanski PM, Waugh TR, Orofino CF, Anzel SH. UCI knee replacement. *Clin Orthop* 1976; 120: 33.
- Ilstrup DM, Combs JJ, Bryan RS, Peterson LFA, Skolnick MD. A statistical evaluation of polycentric total knee arthroplasties. *Clin Orthop* 1976; 120: 18.
- Insall JN, Ranawat CS, Aglietti P, Shine J. A comparison of four models of total knee replacement prostheses. *J Bone Joint Surg* 1976; 58A (6): 754.
- Kettelkamp DB, Thompson C. Development of a knee scoring scale. *Clin Orthop* 1975; 107.
- Marmor L. The modular (Marmor) knee. Case report with a minimum follow-up of 2 years. *Clin Orthop* 1976; 120: 86.
- Tew M, Waugh W. Total replacement of the knee. *J Bone Joint Surg* 1979; 61B: 225–8.
- Weinfeld MS. Proposed rating system for evaluation of knee arthroplasties. *Surg Clin N Amer* 1969; 49 (4).
- Wilson FC, Venters GC. Results of knee replacement with the Wallidius prosthesis. An interim report. *Clin Orthop* 1976; 120: 39.
- Zacharine E. Osteoarthritis treated with indomethacin. *North Med* 1966; 75: 384–5.
- Jonsson GT. Compartmental arthroplasty for gonarthrosis. *Acta Orthop Scand (suppl.)* 1981; 52: 193.