

견관절 Scoring System의 비교연구

중앙대학교 의과대학 정형외과학교실, *동국대학교 공과대학 산업공학과

태석기 · 조성구* · 정영복 · 진희재 · 김종원

— Abstract —

A Comparative Study of the Shoulder Scoring Systems

Suk-Kee Tae, M.D., Sung Koo Cho, Ph.D., Young Bok Jung, M.D., Hui Jae Jin, M.D.
and Jong Won Kim, M.D.

Department of Orthopedic Surgery, College of Medicine, Chung-Ang University
**Department of Industrial Engineering, College of Engineering, Dongguk University*

Aim : To evaluate validity and responsiveness of four shoulder scoring systems.

Material and Method: Twenty-five cases of shoulder instability(22 traumatic, 3 non-traumatic) and twenty-three cases of rotator cuff tear(12 small or medium, 10 large or massive) treated surgically were evaluated with the Shoulder Function Score of the University of Pennsylvania(Penn FS), Constant Score, UCLA Shoulder Rating Scale and Simple Shoulder Test(SST), preoperatively and at final follow-up. The average follow-up was 16.0 months in instability group and 17.5 months in rotator cuff tear group. Using the SPSS program, Pearson linear correlation coefficient(PLCC) between the scores were calculated. And to assess the construct validity, PLCC between patients' satisfaction and the scores were also calculated. Responsiveness was measured by the standardized response mean(SRM).

Result : In instability group, correlation between the scoring systems was low preoperatively except between Constant and SST, but high after operation. Patients' satisfaction with the scores showed low PLCC preoperatively, but high PLCC postoperatively. SRM was high in PENN and UCLA, but when the satisfaction segment of the score was eliminated from UCLA, the SRM was the lowest. In rotator cuff tear group, there was high correlation between the scores not only preoperatively but postoperatively. And the patients' satisfaction matched well with the scores. SRM was particularly high in UCLA and SST. Even when satisfaction segment was eliminated from UCLA, the SRM was still the highest.

Conclusion : Evaluation by the 4 scoring systems investigated in the study showed less consistency in instability than rotator cuff tear in terms of correlation and validity. Responsiveness was generally higher in rotator cuff tear

※통신저자 : 태 석 기

서울특별시 용산구 한강로 3가

중앙대학교 의과대학 정형외과학교실

Tel : 02)748-9835, Fax : 02)793-6634, E-mail : skt97@hananet.net

* 본 논문은 2000년도 중앙대학교 교내 연구비 지원에 의해 작성되었음.

* 본 논문의 요지는 2001년도 대한정형외과 추계학술대회에서 구연되었음

group than in instability group except for Pennsylvania Shoulder Function Score. Therefore it is construed that use of any among the four scoring systems doesn't make difference in evaluation of rotator cuff lesions. However in instability group, care is needed because different result may be obtained according to the selection of a scoring system.

Key Words : Shoulder scoring system, Shoulder instability, Rotator cuff tear

서 론

연구대상 및 방법

견관절을 침범하는 질환은 매우 다양하나 치료 결과의 판정기준은 다른 관절에 비하여 통일된 기준이 없는 실정이다. 현재 흔히 사용되고 있는 견관절 질환에 대한 평가기준은 Neer's Rating System¹⁴⁾, UCLA Shoulder Scoring System⁵⁾, Rowe's Rating Scale¹⁷⁾, Hospital for Special Surgery Shoulder Rating Scale¹⁾, Constant's Scoring System⁴⁾, American Shoulder and Elbow Surgeons' Form¹⁵⁾ 및 Simple Shoulder Test¹³⁾ 등 여러가지가 있다.

일반적으로 임상결과 판정을 위한 기준은 타당성(Validity), 재생성(reproducibility), 내적 일관성(internal consistency), 변화에 대한 반응도(responsiveness to change) 등을 기준으로 판단되어야 한다¹⁸⁾. 그러나 동일한 견관절 질환에서 다른 평가기준의 적용이 어떠한 차이를 보이는가에 대하여는 연구^{2,16)}는 충분하지 않다. 또한 위의 여러 기준은 한국인과는 생활관습 및 문화가 다른 서양인을 기준으로 만들어진 것이므로 한국인에 그대로 적용하는 것이 타당한지에 대한 검증은 되어 있지 않다.

본 연구의 목적은 견관절의 대표적 질환인 불안정증 회전근개 파열 환자의 수술 예들을 대상으로 University of Pennsylvania Shoulder Function Score¹⁰⁾, Constant's Score⁴⁾, UCLA Shoulder Rating Scale⁵⁾ 및 Simple Shoulder Test¹³⁾의 네 가지 기준으로 평가한 결과의 변화를 비교하여 각 척도 사이의 상관관계, 척도의 구조적 타당성(construct validity) 및 반응도(responsiveness)를 알아보는 데 있다.

1998년 1월부터 2000년 8월 사이에 수술적으로 치료한 견관절 전방 불안정증 25례 및 회전근개 파열 환자 23례를 대상으로 하였다. 불안정증 환자는 남자 22례, 여자 3례로서 평균연령은 25.3세(범위 16~35세)이었다. 25중 22례는 외상성 불안정, 3례는 다방향성 불안정이었다. 회전근개 파열 환자는 남자 11례, 여자 12례로서 평균연령은 58세(범위 23~78세)이었으며 12례는 소형 및 중형, 나머지 11례는 대형 이상이었다. 추시기간은 불안정군에서 평균 16개월(범위 12~26개월), 회전근개 파열군에서 평균 17.5개월(범위 12~21개월)이었다.

모든 환자는 수술 전 및 최종 추시점에 University of Pennsylvania Shoulder Function Score(이하 Penn FS로 표기), Constant's Score(이하 Constant로 표기), UCLA Score(이하 UCLA로 표기) 및 Simple Shoulder Test(이하 SST로 표기)의 네 가지 척도를 이용한 평가를 시행하였다. 먼저 척도간의 상관관계를 알기 위하여 최초, 최종 점수간의 Pearson 선형상관계수(Pearson linear correlation coefficient; 이하 PLCC로 표기)를 구하였다. 또한 척도의 구조적 타당성의 간접적인 측정을 위하여 환자의 주관적 만족도와 각 척도간의 PLCC를 구하였다. 척도의 반응도는 표준반응평균(standardized response mean; 이하 SRM으로 표기)¹¹⁾으로서 구하였다. SRM은 점수변화의 평균치를 변화점수의 표준편차로 나눈 것이다.

결 과

1) 불안정증군에서의 결과

각 척도별 최초 점수간 상관분석을 시행한 결과 Pearson 선형상관계수는 Constant와 SST간은

Table 1. Descriptive stastics in instability group.

	N	Average	Standard deviation	Variance
PENN-I	25	69.45	13.93	193.945
PENN-F	25	92.85	7.09	50.239
CONSTANT-I	25	82.85	7.41	54.871
CONSTANT-F	25	89.65	7.03	49.397
UCLA-I	25	27.70	1.17	1.379
UCLA-F	25	33.90	1.77	3.147
SST-I	25	10.30	1.22	1.484
SST-F	25	11.50	1.24	1.526
Satisfaction-I	25	4.20	1.24	1.537
Satisfaction-F	25	9.50	0.76	0.579
UCLA-I(without satisfaction)	25	27.70	1.17	1.379
UCLA-F(without satisfaction)	25	28.90	1.77	3.147

I : Initial score

F : Final score

Table 2. Correlation analysis of the initial scores in instability group.

		PENN-I	CONST-I	UCLA-I	SST-I	Satisfaction-I
PLCC	PENN-I	1.000	.426	-.075	.503*	.074
	Significance probability (two sided)	.	.061	.753	.024	.757
	N	25	25	25	25	25
CONSTANT-I	PLCC	.426	1.000	.031	.758**	-.169
	Significance probability (two sided)	.061	.	.897	.000	.478
	N	25	25	25	25	25
UCLA-I	PLCC	-.075	.031	1.000	.213	.188
	Significance probability (two sided)	.753	.897	.	.366	.427
	N	25	25	25	25	25
SST-I	PLCC	.503*	.758**	.213	1.000	.202
	Significance probability (two sided)	.024	.000	.366	.	.393
	N	25	25	25	25	25
Satisfaction-I	PLCC	.074	-.169	.188	.202	1.000
	Significance probability (two sided)	.757	.478	.427	.393	.
	N	25	25	25	25	25

PLCC : Pearson Linear Correlation Coefficient

* Correlation coefficient significant at P= 0.05(two sided)

** Correlation coefficient significant at P= 0.01(two sided)

0.758로서 매우 유의하였으며 Penn과 SST간은 0.503으로서 유의하였다. 그러나 그 외에는 유의한 상관관계가 보이지 않았다. 또한 주관적 만족도와 척도와의 상관관계도 낮았다(Table 1, 2). 최종점수의 상관분석에서는 모든 척도간에 높은 상관관계를 보였으며 주관적 만족도와 각 척도의

Table 3. Correlation analysis of the final scores in instability group.

		PENN-F	CONST-F	UCLA-F	SST-F	Satisfaction-F
PENN-F	PLCC	1.000	.765**	.698**	.736**	.659**
	Significance probability (two sided)	.	.000	.001	.000	.002
	N	25	25	25	25	25
CONSTANT-F	PLCC	.765**	1.000	.719**	.761**	.635**
	Significance probability (two sided)	.000	.	.000	.000	.003
	N	25	25	25	25	25
UCLA-F	PLCC	.698**	.719**	1.000	.744**	.819**
	Significance probability (two sided)	.001	.000	.	.000	.000
	N	25	25	25	25	25
SST-F	PLCC	.736**	.761**	.744**	1.000	.616**
	Significance probability (two sided)	.000	.000	.000	.	.004
	N	25	25	25	25	25
Satisfaction-F	PLCC	.659**	.635**	.819**	.616**	1.000
	Significance probability (two sided)	.002	.003	.000	.004	.
	N	25	25	25	25	25

PLCC : Pearson Linear Correlation Coefficient

** correlation coefficient significant at P= 0.01(two sided)

Table 4. Responsiveness in instability group.

Measure	Standardized Response Mean
PENN	3.3004
CONSTANT	0.9673
UCLA	3.5028
UCLA(without satisfaction)	0.6780
SST	0.8064

Standardized Response Mean =

$$\frac{\text{Mean change of score}}{\text{Standard deviation of score changes}}$$

점수도 거의 비슷한 정도의 상관관계를 보였다 (Table 1, 3).

Standardized response mean (SRM)으로서 측정된 반응도는 Penn 3.30, UCLA 3.50. 으로서 매우 높았으나 UCLA에서 만족도를 제거할 경우는 0.68로서 가장 낮은 반응도를 보였다 (Table 4).

2) 회전근개 파열군에서의 결과

최초 점수간의 상관분석상 모든 척도간에 매우 유의한 상관관계를 보였으며 특히 Penn과 SST 간의 상관관계가 높게 나타났다. 주관적 만족도와 척도와의 상관관계도 높았다 (Table 5, 6). 최종 점수간의 상관분석에서도 모든 척도들이 매우 유의한 상관관계를 보여 회전근개 파열에서는 최초 및 최종 점수간의 일관성이 높은 것으로 평가되었다. 또한 주관적 만족도와 척도와의 상관관계도 높았다 (Table 5, 7).

SRM은 UCLA가 3.73으로서 가장 높게 나타났으며 SST, Constant, Penn의 순이었다. 또한 UCLA에서 만족도를 제외한 경우에도 SRM은 2.61로서 다른 척도보다 높게 나타났다 (Table 8).

고 찰

임상의학에서 결과 측정방법은 치료의 효과를

Table 5. Descriptive stastics in rotator cuff group.

	N	Average	Standard deviation	Variance
PENN-I	23	39.75	20.23	409.145
PENN-F	23	68.35	18.65	347.924
CONSTANT-I	23	34.55	11.80	139.313
CONSTANT-F	23	67.30	17.56	308.432
UCLA-I	23	12.90	5.41	29.253
UCLA-F	23	29.60	4.48	20.042
SST-I	23	2.75	1.55	2.408
SST-F	23	7.00	2.26	5.116
Satisfaction-I	23	2.05	1.19	1.418
Satisfaction-F	23	7.04	1.39	1.937

I : Initial score

F : Final score

Table 6. Correlation analysis of the initial scores in rotator cuff group.

		PENN-I	CONST-I	UCLA-I	SST-I	Satisfaction-I
PENN-I	PLCC	1.000	.703**	.655**	.828**	.483*
	Significance probability (two sided)	.	.001	.002	.000	.031
	N	23	23	23	23	23
CONSTANT-I	PLCC	.703**	1.000	.773**	.749**	.552*
	Significance probability (two sided)	.001	.	.000	.000	.012
	N	23	23	23	23	23
UCLA-I	PLCC	.655**	.773**	1.000	.774**	.597**
	Significance probability (two sided)	.002	.000	.	.000	.005
	N	23	23	23	23	23
SST-I	PLCC	.828**	.749**	.774**	1.000	.520*
	Significance probability (two sided)	.000	.000	.000	.	.019
	N	23	23	23	23	23
Satisfaction-I	PLCC	.483*	.552*	.597**	.520*	1.000
	Significance probability (two sided)	.031	.012	.005	.019	.
	N	23	23	23	23	23

PLCC : Pearson Linear Correlation Coefficient

* Correlation coefficient significant at P= 0.01(two sided)

** Correlation coefficient significant at P= 0.05(two sided)

잘 알 수 있게 할 뿐 아니라 의사들간의 의사소통 수단이다. 결과 측정방법의 선택에 따라 달라서 동일한 환자에서의 결과가 다르게 판정될 수도 있다⁶⁾. 따라서 어떠한 척도가 치료 결과의 측정에

좋은 척도인가를 아는 것은 매우 중요하다.

견관절 평가 척도는 임상연구의 목적, 전통 등에 따라 여러 가지가 사용되며 임상연구의 일부분으로서 평가 척도에 대한 비교가 행하여지고 있다⁸⁾.

Table 7. Correlation analysis of the final scores in rotator cuff group.

		PENN-F	CONST-F	UCLA-F	SST-F	Satisfaction-F
PENN-F	PLCC	1.000	.892**	.848**	.819**	.779**
	Significance probability (two sided)	.	.000	.000	.000	.000
	N	23	23	23	23	23
CONSTANT-F	PLCC	.892**	1.000	.871**	.893**	.682**
	Significance probability (two sided)	.000	.	.000	.000	.001
	N	23	23	23	23	23
UCLA-F	PLCC	.848**	.871**	1.000	.881**	.686**
	Significance probability (two sided)	.000	.000	.	.000	.001
	N	23	23	23	23	23
SST-F	PLCC	.819**	.893**	.881**	1.000	.562**
	Significance probability (two sided)	.000	.000	.000	.	.010
	N	23	23	23	23	23
Satisfaction-F	PLCC	.779**	.682**	.686**	.562**	1.000
	Significance probability (two sided)	.000	.001	.001	.010	.
	N	23	23	23	23	20

PLCC : Pearson Linear Correlation Coefficient

** correlation coefficient significant at P= 0.01(two sided)

Table 8. Responsiveness in rotator cuff group.

Measure	Standardized Response Mean
PENN	1.5335
CONSTANT	1.8650
UCLA	3.7277
UCLA(without satisfaction)	2.6116
SST	2.2345

$$\text{Standardized Response Mean} = \frac{\text{Mean change of score}}{\text{Standard deviation of score changes}}$$

평가척도는 single-parameter assessment⁹⁾와 점수 척도(point scale)에 의한 것¹⁹⁾에 의한 것으로 구분된다. 요즈음은 점수 척도에 의한 것이 많이 고안, 사용되고 있는데⁴⁾ 이는 관찰자간 차이 및 측정하는 매개변수(parameter)에 따른 자료의 왜곡을 줄이는 것을 목적으로 한다¹⁶⁾. 또한 많은 scoring system의 고안 및 사용에 따라 각 기준의 타당성(validity), 반복성(repeatibility) 및 민감성(sensitivity) 등의 중요성에 대한 관심

도 증대되고 있다¹⁵⁾. 한편 각 질환별로 특정 scoring system이 결과를 잘 반영하는가에 대한 연구도 이루어 지고 있다³⁾. 그러나 국내에서는 아직 이에 대한 연구보고는 없으며 주로 외국의 평가기준을 그대로 사용하고 있는 실정이다.

본 연구에 사용한 네 척도 중 Penn FS와 SST와 같은 설문 척도는 환자 자신이 작성할 수 있으므로 의사가 직접 환자를 만나지 않아도 측정이 가능하여 우편으로 추시 자료를 얻을 수도 있어 추시율을 높일 수 있다¹²⁾. 그러나 Constant score와 UCLA score는 근력 및 운동범위 측정을 위하여 의사의 진찰을 요한다.

원칙적으로 좋은 결과판정 척도는 타당성(vadility)이 있어야 하며 변화를 잘 반영하여야 한다. 그러나 타당성은 상태를 반영하는 gold standard(예를 들면 신장, 체중과 같이 수치화할 수 있는 것)가 없는 경우에는 알기가 쉽지 않다. 본 연구에서는 척도의 구조적 타당성(construct validity)을 알기 위하여 두가지 방법을 사용하였다. 첫째는 각 척도간의 상관관계를 구한

것이다. 각 척도를 사용한 점수간의 상관관계가 높으며 일관성이 있다면 이는 척도의 타당성이 높을 가능성이 큰 것을 의미한다. 반대로 점수간의 상관관계가 높지 않고 일관성이 없다면 이는 일부 또는 전체 척도의 타당성이 낮음을 의미한다. 둘째는 환자의 주관적 만족도와 점수간의 상관관계를 구한 것이다. 일반적으로 치료자는 자신이 치료한 환자에서의 결과를 좋게 보려는 경향이 있으며 임상검사변수는 신뢰성에 문제가 있을 수 있기 때문에 임상결과의 판정, 특히 기능적 결과가 중요한 정형외과 질환의 결과 판정에서는 환자 자신의 만족도가 중요하다고 볼 수 있다^{4,6)}. 그 결과 불안정군에서는 척도별로 점수의 일관성이 결여되었으며 만족도와와의 상관관계 검증에서도 비슷한 양상을 보여 일부 또는 전 척도의 타당성에 문제가 있는 것으로 보인다. 또한 변화를 반영하는 정도인 반응도를 SRM으로 검증한 결과 Penn FS와 UCLA에서 높게 나왔으나 UCLA에서 만족도(35점 중 5점) 부분을 제거할 경우는 가장 낮게 나와 Penn FS를 제외한 나머지는 불안정군에서 실제의 상태변화를 적절히 반영하지 못하는 것으로 보인다. 반면에 회전근개 파열군에서는 최초 및 최종 점수간의 상관관계 및 만족도와 각 척도간의 상관관계 검증 결과 일관성이 있는 것으로 보였다. 이는 네 척도의 타당성이 공히 높을 가능성이 큰 것을 의미한다. 또한 SRM은 UCLA에서 가장 높았으며 만족도 부분을 제거한 경우에도 다른 척도보다 높게 나왔다. 이는 회전근개 파열에서는 결과 판정에 UCLA Rating Scale이 매우 우수한 기준임을 시사하는 것이다.

본 연구는 단 1회의 측정에 의하여 얻은 점수를 토대로 상관관계 및 반응도 등을 구한 것으로서 점수의 신뢰성(reliability)에 문제가 있을 가능성을 배제할 수 없으나 회전근개 파열군에서 사용된 네 가지 척도가 공히 일관성을 보였으며 높은 반응도를 보인 것은 의미 있는 결과로 생각된다. 즉, 회전근개 파열에 대한 치료결과의 판정에는 네가지의 척도중 어느 것을 사용하더라도 무방한 것으로 보인다. 그러나 불안정증군에서는 네가지 척도간 상관관계 검증에서 점수간의 일관성이 결여되고 반응도에도 Penn FS를 제외하고는 문제가 있는 것으로 보여 불안정증 환자의 치료 결과

판정시 적절한 척도의 선정에 신중을 기하여야 할 것으로 사료된다.

요약 및 결론

견관절 불안정증 및 회전근개 파열 환자에서 Pennsylvania Shoulder Scoring System, UCLA Shoulder Rating Scale, Constant Score 및 Simple Shoulder Test로 수술 전 후에 측정된 점수의 척도간 상관관계, 척도의 타당성 및 반응도를 검증한 결과 불안정증에서는 각 척도를 사용한 점수의 일관성이 낮은 것으로 나타났다. 회전근개 파열에서는 동일한 방법으로 검증한 결과 일관성이 높았다. 회전근개 파열 환자에서는 검증된 네가지 척도 중 어느 것을 사용하여도 무방할 것으로 보이거나 불안정증에서는 평가 척도의 선정에 주의를 요할 것으로 사료된다.

REFERENCES

- 1) **Altcheck DW, Warren RF and Wickiewicz TL** : Arthroscopic acromioplasty. Technique and results. *J Bone Joint Surg*, 72-A: 1198-1207, 1990.
- 2) **Beaton DE, Richards RR** : Measuring function of the shoulder. A cross-sectional comparison of five questionnaires. *J Bone Joint Surg*, 78-A: 882-890, 1996.
- 3) **Conboy VB, Morris RW and Carr AJ** : An evaluation of the Constant-Murly shoulder assessment. *J Bone Joint Surg*, 78-A: 229-232, 1996.
- 4) **Constant CR and Murley AHG** : A clinical method of functional assessment of the shoulder. *Clin Orthop*, 214: 160-164, 1987.
- 5) **Ellman H, Haker G and Bayer M** : Repair of the rotator cuff: end-result study of factors influencing reconstruction. *J Bone Joint Surg*, 68-A: 1136-1144, 1988.
- 6) **Fuchs B, Jost B and Gerber C** : Posterior-inferior capsular shift for the treatment of recurrent, voluntary posterior subluxation of the shoulder. *J Bone Joint Surg*, 82A: 16-25, 2000.
- 7) **Gazielly DF, Gleyze P and Montagnon C** : Functional and anatomical results after rotator cuff repair. *Clin Orthop*, 304: 43-53, 1994.

- 8) **Keller RB** : How outcome research should be done. In *The shoulder. A balance of mobility and stability.* Matsen FA et al(eds), Rosemont, American Academy of Orthopedic Surgeons, 497-499, 1993.
- 9) **Kessel L and Barley I** : Prosthetic replacement of shoulder joint. *J R Soc Med*, 72: 748-755, 1979.
- 10) **Leggin BG and Iannotti JP** : Shoulder outcome measurement. In: *Disorders of the shoulder: diagnosis and management.* Iannotti JP and Williams GR(eds.), Lippincott, Williams and Wilkins, Philadelphia, 1999.
- 11) **Liang MH, Fossel AH and Larson MG** : Comparison of five health status instruments for orthopedic evaluation. *Med Care*, 28: 632-642, 1990.
- 12) **L'Insanata JC, Warren RF, Cohen SB, Altchek DW and Peterson MGE** : A self-administered questionnaire for assessment of symptoms and function of the shoulder. *J Bone Joint Surg*, 79-A: 738-748, 1997.
- 13) **Lippit S, Harryman DT and Matsen FA** : A practical tool for evaluating function: The simple shoulder test. In: Matsen FA et al(eds.) *The shoulder: a balance of mobility and stability.* Rosemont. American Academy of Orthopedic Surgeons, 1993.
- 14) **Neer CS II** : Anterior acromioplasty for the chronic impingement syndrome in the shoulder: a preliminary report. *J Bone Joint Surg*, 54-A: 41-50, 1972.
- 15) **Richards RR, An K-N and Bigliani LU** : A standardized method for assessment of shoulder function. *J Shoulder Elbow Surg*, 3: 347-352, 1994.
- 16) **Romeo AA, Bach BR, Jr and O'Halloran KL** : Scoring system for shoulder conditions. *Am J Sports Med*, 24: 472-476, 1996.
- 17) **Rowe CR, Patel D and Southmayd WW** : The Bankart procedure: A long-term end-result study. *J Bone Joint Surg*, 60-A: 1-16, 1978.
- 18) **Ware JE** : Methodological considerations in the selection of health status assessment procedures. In Wenger NK, et al(eds): *Assessment of Quality of Life in Clinical Trials on Cardiovascular Therapies.* New York: Le Jac, pp 87-111, 1984.
- 19) **Wolfgang GL** : Surgical repair of tears of the rotator cuff of the shoulder. *J Bone Joint Surg*, 56-A: 14-20, 1974.