Osteoarthritis and Cartilage (2007) **15,** 1397–1402

doi:10.1016/j.joca.2007.05.005

Osteoarthritis and Cartilage

International Cartilage Repair Society

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

International Cartilage Repair Society (ICRS) and Oswestry macroscopic cartilage evaluation scores validated for use in Autologous Chondrocyte Implantation (ACI) and microfracture[1]

M. P. J. van den Borne M.D., Resident Orthopaedic Surgery†, N. J. H. Raijmakers M.Sc., Orthopaedic Clinical Research Manager†, J. Vanlauwe M.D., Orthopaedic Surgeon‡, J. Victor M.D., Orthopaedic Surgeon§, S. N. de Jong M.D., Arthroscopy Fellow†, J. Bellemans M.D., Ph.D., Orthopaedic Surgeon‡ and D. B. F. Saris M.D., Ph.D., Orthopaedic Surgeon†*

† *Department of Orthopaedics, University Medical Center Utrecht, The Netherlands*

‡ *Department of Orthopaedics, University Hospital Leuven, Belgium*

§ *Department of Orthopaedics, Medical Centre Sint-Lukas, Brugge, Belgium*

Summary

Objective: For young patients with cartilage defects, the emergence of clinically applicable cell therapy for biological joint reconstruction is an appealing prospect. Acceptation of this method as a means of standard care requires proof of being reproducible, having long-lasting mechanical integrity, and having a good clinical outcome. This study evaluates the reliability of the International Cartilage Repair Society (ICRS) score and the Oswestry Arthroscopy Score (OAS) in the assessment of regenerative cartilage repair.

Method: A total of 101 macroscopic images of cartilage repair were made during arthroscopy 12 months post-treatment of either Autologous Chondrocyte Implantation (ACI) or microfracture. These images were examined by seven independent observers with differing levels of experience. The ICRS and OAS scores were randomly presented twice at a 4-week interval. All observers stated their predicted outcome according to actual treatment and defect size.

Results: ICRS and OAS scores showed both good inter- and intra observer reliability (0.62 and 0.56 for ICRS; 0.73 and 0.65 for OAS, respectively). Internal consistency (Cronbach's alpha) was satisfactory for research purposes (0.79 and 0.74, respectively). Correlation (equivalence concordance) between both scoring systems was excellent ($r = 0.94$). All observers were inconsistent in predicting actual treatment. Test–re test reliability of estimated defect size and its correlation to true defect size were poor. These results were also applicable to the subanalyses of the experience of the observer and the quality of imaging.

Conclusion: The ICRS and OAS are reliable and relevant scores that are now both validated for macroscopic evaluation of cartilage repair as a research tool.

Key words: Validation, Cartilage repair, ICRS, OAS, ACI, Microfracturing.

Introduction

There is an increase in the number of published reports on the clinical outcome of cartilage repair. Different treatment modalities are being designed, changed, trialled and published. However, controversy about the superiority of the different treatment options still exists[1] and the methodological quality of studies is generally low[2]. In comparing these different modalities, a uniform and reliable scoring system for assessing cartilage repair is desirable, and should be the cornerstone of evidenced based cartilage repair surgery repair selection. Different methods of evaluation are available, either structural or clinical, biopsy or surface evaluation, patient or doctor administrated. Assessment of structural outcome of cartilage repair can be done by Magnetic Resonance Imaging (MRI), histology and macroscopic evaluation through arthroscopy. Different evaluation systems for MRI exist[3–5]; one of the most important being that of Marlovits[6] who in 2004 introduced a definition of pertinent parameters for evaluating cartilage repair. However, scan protocols, hardware and software capabilities rapidly change and outcome tools are thus inconsistent. Different histological scoring systems in cartilage repair exist such as the comprehensive O'Driscoll[7–9], the simple Pineda scale[7,10], the Bern score[11] and the International Cartilage Repair Society (ICRS) Visual Histological Assessment Scale[12]. Some of these have been validated and various modifications have been applied, but still no general consensus exists.

For macroscopic evaluation of cartilage repair two scoring systems are available; the ICRS and the Oswestry Arthroscopy Score (OAS). That of the ICRS score was designed by Brittberg and Peterson[13,14] (Table I); The OsCell group of Oswestry developed the OAS[15] (Table II), in an attempt to simplify and focus the scoring system on clinical needs.

[1]Investigations were performed in the Department of Orthopaedics at the University Medical Center Utrecht, The Netherlands.

*Address correspondence and reprint requests to: Dr D. B. F. Saris, M.D., Ph.D., Orthopaedic Surgeon, Department of Orthopaedics, University Medical Center Utrecht, P.O. Box 85500, 3508 GA Utrecht, The Netherlands. Tel: 31-30-2506971; Fax: 31-30-2510638; E-mail: d.saris@umcutrecht.nl

Received 23 December 2006; revision accepted 1 May 2007.

Table I
ICRS macroscopic evaluation of cartilage repair

Cartilage repair assessment ICRS	Points
Degree of defect repair	
In level with surrounding cartilage	4
75% repair of defect depth	3
50% repair of defect depth	2
25% repair of defect depth	1
0% repair of defect depth	0
Integration to border zone	
Complete integration with surrounding cartilage	4
Demarcating border < 1 mm	3
3/4th of graft integrated, 1/4th with a notable border > 1 mm width	2
1/2 of graft integrated with surrounding cartilage, 1/2 with a notable border > 1 mm	1
From no contact to 1/4th of graft integrated with surrounding cartilage	0
Macroscopic appearance	
Intact smooth surface	4
Fibrillated surface	3
Small, scattered fissures or cracs	2
Several, small or few but large fissures	1
Total degeneration of grafted area	0
Overall repair assessment	
Grade I: normal	12
Grade II: nearly normal	11–8
Grade III: abnormal	7–4
Grade IV: severely abnormal	3–1

These two scoring systems were specifically designed to evaluate the macroscopic outcome of cartilage repair in an attempt to simplify and focus the scoring system on clinical needs.

To date, only Smith *et al.*[15] have made an attempt to validate Autologous Chondrocyte Implantation (ACI) at a time interval of 2 months by assessment of five videos by six orthopedic surgeons. They showed the ICRS and the OAS to be effective tools in the evaluation of cartilage repair.

Table II
OAS (Oswestry Arthroscopy Score)

OAS	Points
Graft level with surrounding cartilage	
Level	2
Raised	1
Below	0
Integration with surrounding cartilage	
Complete	2
Minor disruption (<25% of area)	1
Major disruption (>25% of area)	0
Appearance of surface	
Smooth	2
Fine fronds	1
Severe fronds/fibrillation	0
Color of graft	
Pearly, hyaline-like	2
White	1
Yellow bone	0
Stiffness on probing	
Normal compared to adjacent cartilage	2
Softer	1
Very soft/hard	0
Total	0–10

In evaluating the results of macroscopic cartilage repair, a valid reliable scoring system is needed. It should be an unfailing indicator of the pathophysiological condition and should have good validity. Further it should be applicable to various treatment modalities and the outcome should have bearing on functional and clinical needs. Therefore we set out to test the reliability and validity of the International Cartilage Repair Score (ICRS) and the OAS used for macroscopic evaluation of cartilage repair after ACI and microfracturing in an extensive evaluation of 101 macroscopic images by seven observers using both scores at two time points.

Materials and methods

MACROSCOPIC IMAGES

Macroscopic evaluation of cartilage repair was done, using images (videos and prints) made during arthroscopy 12 months after previous cartilage repair surgery. These procedures were an essential part of a prospective multicenter randomized controlled trial, comparing ACI using a periosteal flap vs microfracture treatment in the repair of symptomatic cartilage defects of the knee (*paper submitted*). A total of 118 patients were included in the trial; of these, 107 patients underwent an arthroscopic procedure at 12 months post-cartilage repair surgery and images of the arthroscopic exam were collected. A total of 101 macroscopic images were evaluated, 52 cases of microfracture (21 prints and 31 videos) and 49 of ACI (16 prints and 33 videos). Images from six patients were missing, lost either during distribution and/or with problems in saving data to disk or network.

All of the 101 videos and prints were collected and edited in a central electronic database with the purpose only to show the cartilage repair site. A Filemaker Pro format was then designed to show both the images and score systems on the same screen (FileMaker Inc., Santa Clara, CA, USA). An example of differences in outcome is shown in Fig. 1.

EXAMINATION

The 101 edited images of the arthroscopic procedures were presented to seven observers; four orthopedic surgeons with extensive experience in cartilage surgery, one arthroscopy fellow, one orthopedic surgery resident and one clinical research manager with a non-medical background. Each observer was asked to score the images by using the central electronic database. In addition, the observers were asked to judge the quality of the images and to estimate the defect size. The quality of imaging was judged by poor quality (0 points), moderate quality (1 point) or good quality (2 points). The overall quality of the images was the sum of the judgment of all of the observers and was interpreted as follows: poor quality (0–4 points), moderate quality (5–9 points) or good quality (10–14 points). The defect size was estimated in mm^2. All images were blinded by a unique randomized number and were presented in a random order. After a period of 4 weeks, the images were presented in a different random order and were judged for a second time by all of the observers. For each image, the observers were also asked to specify which repair method (microfracture or ACI) they thought was used.

Fig. 1. Illustration of three typical examples of freezed fragments of video films after cartilage repair/regeneration. The quality of cartilage regeneration/repair was judged to be good in "A", moderate in "B" and poor in "C".

STATISTICAL ANALYSES

Reliability

Intra- and inter observer reliability were assessed by calculating the Intra class Correlation Coefficient (ICC) for both measurement systems and for the estimated defect size. As described by Fleiss[16], the test–re test reliability (concordance) is considered good if the ICC is 0.4–0.75 and excellent if the ICC $>$ 0.75.

Reliability also encompasses the internal consistency of a scale, usually measured as Cronbach's alpha (α), which contains the degree of correlation among items. Internal consistency of each scale was estimated using the Cronbach's α coefficient. For comparing groups, α values of 0.7–0.8 are regarded as satisfactory. For the clinical application, much higher α values are needed, with a minimum of 0.90[17].

Validity

The Pearson correlation coefficient was used for testing equivalence correlation of both scoring systems and correlation between the mean estimated defect size (mean of the first and second evaluation).

DEFECT SIZE EVALUATION AND TREATMENT GUESS

The ICC was also determined for the estimated defect size and comparison to true defect size. Kappa values were determined to correlate the treatment prediction during the first and second evaluation, and interpreted according to the guidelines described by Landis and Koch[18]: $<$0.00 as poor, 0.00–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial and 0.81–1.00 as almost perfect.

Kappa values were also determined to correlate the treatment guess to the performed treatment, and interpreted as described above.

For the sub-analyses, all of the data were divided into groups according to the quality of images and the experience of the observers in order to determine the effect of these parameters.

Statistical analyses were performed using the SPSS 12.0 software for Windows (SPSS Inc., Chicago, IL, USA).

Results

To test the validity and reliability of both ICRS and OAS scores, seven observers judged 101 macroscopic images of cartilage repair (12 months after arthroscopic surgery), twice, with an interval of 4 weeks. Results are shown in Table III.

RELIABILITY

Inter-observer reliability as determined by ICC was 0.62 for ICRS and 0.56 for OAS. Test–re test reliability (intra observer reliability) was 0.73 [standard deviation (SD) 0.05] for ICRS and 0.65 (SD 0.08) for OAS, respectively. Both have good intra- and inter observer reliability according to the criteria described by Fleiss[16]. Internal consistency (Cronbach's α) was 0.79 for ICRS and 0.74 for OAS, respectively. The reliability for both scores was judged to be

Table III
Results of the intra- and inter observer reliability of the ICRS macroscopic evaluation of cartilage repair and the OAS, of the whole group, and sub-analysis of experienced observer and good quality group

	ICRS	OAS
	ICC mean (SD)	ICC mean (SD)
Whole group		
Intra observer	0.73 (0.05)	0.65 (0.08)
Inter observer		
*t*1	0.62	0.58
*t*2	0.61	0.54
Mean *t*1/*t*2	0.62	0.56
Internal consistency (α)		
*t*1	0.80	0.75
*t*2	0.77	0.73
Mean *t*1/*t*2	0.78	0.71
Experienced observer group		
Intra observer	0.70 (0.04)	0.61 (0.05)
Inter observer		
*t*1	0.62	0.53
*t*2	0.61	0.48
Mean *t*1/*t*2	0.61	0.51
Internal consistency (α)		
*t*1	0.84	0.76
*t*2	0.82	0.77
Mean *t*1/*t*2	0.83	0.76
Good quality group		
Intra observer	0.74 (0.11)	0.65 (0.11)
Inter observer		
*t*1	0.66	0.59
*t*2	0.69	0.58
Mean *t*1/*t*2	0.68	0.58
Internal consistency (α)		
*t*1	0.76	0.72
*t*2	0.76	0.76
Mean *t*1/*t*2	0.76	0.74

*t*1 = first evaluation. *t*2 = second evaluation.

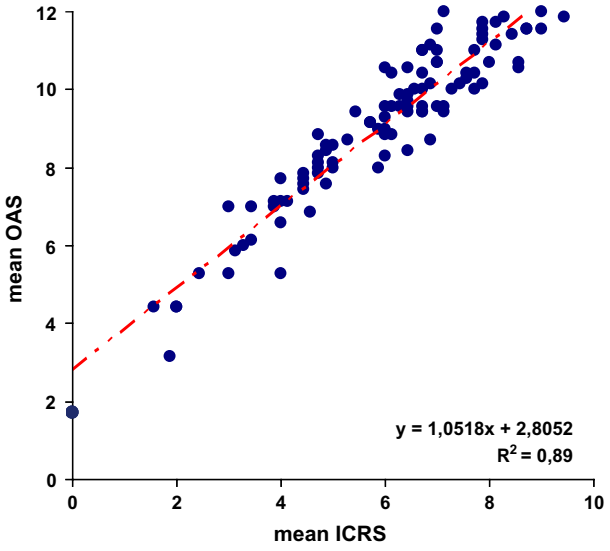

Fig. 2. This graph shows the graphical correlation of the mean ICRS and OAS of all observers, $r = 0.94$. The agreement between both scores is excellent.

satisfactory for research purposes, but not for individual testing[18].

VALIDITY

Equivalence concordance determined by Pearson correlation coefficient specifies the measure of agreement between the scores of ICRS and OAS. This value was 0.94 (Fig. 2), which indicates an excellent correlation.

DEFECT SIZE ESTIMATION AND TREATMENT PREDICTION

The test–re test reliability of the estimated defect size determined by ICC was poor (0.38; SD 0.15). In addition, the correlation between estimated defect size and true size determined by ICC was also poor (0.21; SD 0.09). Test–re test reliability of treatment prediction was fair to moderate with a mean Kappa value of 0.40 (range 0.24–0.55).

The correlation between treatment prediction and the performed treatment was slight with a Kappa value of 0.10. On average, the observers chose the correct treatment in 55.1% of cases (range 39.6–70.3%).

EFFECT OF EXPERIENCED OBSERVER

In order to investigate the influence of the experience of the observer, the results of the four orthopedic surgeons were analyzed separately. The inter-observer reliability determined by ICC was 0.61 for ICRS and 0.51 for OAS, respectively, slightly lower than compared to the values of the whole group (0.62 for ICRS and 0.56 for OAS, respectively). The intra-observer reliability determined by ICC among the orthopedic surgeons was 0.70 (SD 0.04) for ICRS and 0.61 (SD 0.05) for OAS, respectively. These values were also slightly lower than those compared to the whole group (0.73 for ICRS [SD 0.05] and 0.65 for OAS [SD 0.08], respectively).

The test–re test reliability of the estimated defect size by the orthopedic surgeons was 0.36, again slightly lower than compared to the whole group (0.38; SD 0.15). Correlation between the estimated defect size and true size determined by ICC was poor (0.22), but this was slightly better than compared to the whole group (0.21). Test–re test reliability of treatment prediction among the experts showed a mean Kappa value of 0.39.

Internal consistency among the orthopedic surgeons was better than compared to the whole group. Cronbach's α was 0.83 for ICRS and 0.76 for OAS vs 0.78 for ICRS and 0.71 for OAS.

For the orthopedic surgeons, the equivalence reliability between both scores was slightly worse compared to the whole group (0.92 vs 0.94). The treatment prediction compared to the performed treatment was slightly improved (Kappa = 0.14; 56.8% vs 0.10; 55.1%).

The three observers who provided the most correct treatment prediction were compared to the three least reliable observers as determined by Pearson correlation coefficient. These observers chose the correct treatment in 62.5% of cases, compared to 47.4% for the latter. For the observers who predicted the correct treatment most frequently, two were orthopedic surgeons and the third, the clinical research manager with no medical background. The intra observer reliability determined by ICC was 0.73 for ICRS and 0.59 for OAS, thus there were only minor and non-significant differences when compared to the whole group.

EFFECT IMAGE QUALITY

Given that experience does not seem to be of the greatest importance in interpreting reliability of outcome using these scores, the influence of the quality of images was further analyzed. Images were divided in quality groups: 35 images were assessed as *good* quality, 42 as *moderate* quality, and 24 as *poor* quality. The 35 good quality images were only videos. Among these, the inter-observer reliability was slightly improved compared to the whole group; the mean ICC among all seven observers was 0.68 for ICRS and 0.58 for OAS. The test–re test reliability of ICRS and OAS was also slightly better scoring 0.74 and 0.65, respectively. In addition, test–re test reliability of the estimated defect size was slightly better with an ICC value of 0.39. Correlation of the estimated defect size and of its true size was also better compared to the whole group (0.26). However, test–re test reliability of treatment prediction was inferior compared to the whole group (Kappa = 0.36). Internal consistency was slightly differed with Cronbach's α values of 0.76 for ICRS and 0.74 for OAS.

The equivalence reliability between both scores showed some improvement with a Pearson correlation coefficient of 0.96. The treatment prediction was less reliable in the good quality group compared to the whole group with a Kappa value of −0.06.

In summary, there was good agreement (equivalence reliability) between both scores the correlation coefficient determined by Pearson $r = 0.94$. Inter-observer reliability (validity) and intra observer reliability (test–re test reliability) were satisfactory for both scores (range 0.56–0.73). Cronbach's α for both scores also indicated similar internal consistency. Results of estimated defect size and treatment prediction showed poor inter- and intra observer reliability. Sub-analyses relating to the experience of the observers were not greatly improved. Sub-analyses of the quality of images showed a trend toward better consistency between different observers.

Discussion

This study demonstrates that the ICRS macroscopic score and the OAS score are both valid research instruments for

comparing the results of cartilage repair surgery. Intra- and inter observer variability were deemed to be adequate for both ICRS and OAS. Equivalence reliability for both tests was excellent. Internal consistency was satisfactory, but showed that neither tests were adequate for individual testing.

These conclusions remain applicable in the sub-analyses of selected data; based on the experience of the orthopedic surgeons or the quality of the presented images.

Editing of images can cause a bias as only part of the surgical procedure is shown, and also no impression about tactile assessment can be made from a video alone. When using prints it will be even more difficult to judge the repair/regeneration site. A surgeon is expected to have more reliable test results when using the scoring systems during arthroscopy. In this case, the reliability and validity of the tests might be expected to be better.

Smith *et al.*[15] also concluded that both ICRS and OAS are effective tools for macroscopic evaluation of cartilage repair. However, these conclusions were based on only five video images after a single treatment modality (ACI). With a larger group of 101 videos and prints and two different treatments (ACI vs microfracturing) this study confirms that the ICRS score can be considered to be a proper tool and the OAS a reliable alternative.

The study of Smith *et al.*[15] shows higher intra- and inter observer reliability and internal consistency than the current study. Even if the experience of the orthopedic surgeons is compared in both studies; that of Smith shows higher reliability[15]. An explanation can be that the observers in the Smith *et al.*[15] study had only to evaluate five video images instead of a combination of 101 video and still images. The conclusion being that the concentration of the observers is probably less reliable in a more time consuming evaluation (i.e., a greater number of images), and with a small number of images the reproducibility is increased during re test. In addition, the SD will be influenced by the total number of patients.

It became apparent that it was difficult to correctly guess which treatment had been performed based on the images of cartilage repair outcome. Also as had been previously suggested, the estimation of the defect size was extremely irreproducible. The study of Oakley *et al.*[19] reported a variation in accuracy and poor inter-observer reliability of cartilage lesion measurement with the use of conventional methods. In the images used in this study, there were no standardized measurements, which can explain the slightly disappointing results.

Although both scoring systems have their obvious advantages and good reliability, they also have some disadvantages. Smith *et al.*[15] have previously concluded that for the ICRS score, no allowance is made for hypertrophy, however, that is corrected for in the OAS. Hypertrophy is of clinical importance as it can occur after either ACI or microfracturing procedures[20,21].

In addition, the ICRS does not allow for tactile assessment whereas the OAS does. However, the option for ''harder'' cartilage compared to normal cartilage remains absent, with only the option for ''very hard'' cartilage remaining in place. In contrast, there is an option for ''softer'' cartilage as well as there being an option for ''very soft'' cartilage. For example, after microfracture, the repair tissue can sometimes feel harder because of an underlying bone front which shows progress in time and depending on the amount of tissue above it, this can feel hard or very hard.

In conclusion, cartilage repair should be evaluated with use of a scoring system that considers the volume of the defect that becomes filled with repair tissue, the integration of repair tissue with adjacent cartilage, and the macroscopic appearance and biomechanical properties of the repair site[22]. The macroscopic assessment is particularly important in evaluating cartilage repair because it provides information about the quality of the full repair site compared to the incidental histological assessment which only evaluates a biopsy of the repair site.

This study shows that the ICRS and the OAS scoring systems are both useful tools for macroscopic evaluation of cartilage repair for research purposes, but not for individual clinical testing. In all validation tests, the ICRS system scored slightly higher than the OAS score system. Based on this study both scoring systems can be used with confidence, with international standardization of outcome evaluation in mind we feel it merits serious consideration to implement the ICRS score as a validated and reliable outcome for judging the result of articular cartilage regeneration or repair.

Acknowledgements

Dr Saris is supported in part by The Netherlands Organisation for Health Research and Development, Clinical Fellow program. We would like to thank Dr J. Ward for her excellent work in improving the text. Her input as native English speaking medical writer has been a valuable contribution.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.joca.2007.05.005.

References

1. Knutsen G, Engebretsen L, Ludvigsen TC, Drogset JO, Grontvedt T, Solheim E, *et al*. Autologous chondrocyte implantation compared with microfracture in the knee. A randomized trial. J Bone Joint Surg Am 2004;455–64.
2. Jakobsen RB, Engebretsen L, Slauterbeck JR. An analysis of the quality of cartilage repair studies. J Bone Joint Surg Am 2005;2232–9.
3. Alparslan L, Winalski CS, Boutin RD, Minas T. Postoperative magnetic resonance imaging of articular cartilage repair. Semin Musculoskelet Radiol 2001;345–63.
4. Brown WE, Potter HG, Marx RG, Wickiewicz TL, Warren RF. Magnetic resonance imaging appearance of cartilage repair in the knee. Clin Orthop Relat Res 2004;214–23.
5. Winalski CS, Minas T. Evaluation of chondral injuries by magnetic resonance imaging: repair assessments. Tech Sports Med 2000;8:108–19.
6. Marlovits S, Striessnig G, Resinger CT, Aldrian SM, Vecsei V, Imhof H, *et al*. Definition of pertinent parameters for the evaluation of articular cartilage repair tissue with high-resolution magnetic resonance imaging. Eur J Radiol 2004;310–9.
7. Moojen DJ, Saris DB, Auw Yang KG, Dhert WJ, Verbout AJ. The correlation and reproducibility of histological scoring systems in cartilage repair. Tissue Eng 2002;627–34.
8. O'Driscoll SW, Keeley FW, Salter RB. The chondrogenic potential of free autogenous periosteal grafts for biological resurfacing of major full-thickness

defects in joint surfaces under the influence of continuous passive motion. An experimental investigation in the rabbit. J Bone Joint Surg Am 1986;1017–35.
9. O'Driscoll SW, Keeley FW, Salter RB. Durability of regenerated articular cartilage produced by free autogenous periosteal grafts in major full-thickness defects in joint surfaces under the influence of continuous passive motion. A follow-up report at one year. J Bone Joint Surg Am 1988;595–606.
10. Pineda S, Pollack A, Stevenson S, Goldberg V, Caplan A. A semiquantitative scale for histologic grading of articular cartilage repair. Acta Anat (Basel) 1992; 335–40.
11. Grogan SP, Barbero A, Winkelmann V, Rieser F, Fitzsimmons JS, O'Driscoll, *et al*. Visual histological grading system for the evaluation of *in vitro*-generated neocartilage. Tissue Eng 2006;2141–9.
12. Mainil-Varlet P, Aigner T, Brittberg M, Bullough P, Hollander A, Hunziker E, *et al*. Histological assessment of cartilage repair: a report by the Histology Endpoint Committee of the International Cartilage Repair Society (ICRS). J Bone Joint Surg Am 2003;45–57.
13. Peterson L, Minas T, Brittberg M, Nilsson A, Sjogren-Jansson E, Lindahl A. Two- to 9-year outcome after autologous chondrocyte transplantation of the knee. Clin Orthop Relat Res 2000;212–34.
14. Brittberg M, Peterson L. Introduction of an articular cartilage classification. ICRS Newsletter 1998;1:5–8.
15. Smith GD, Taylor J, Almqvist KF, Erggelet C, Knutsen G, Garcia PM, *et al*. Arthroscopic assessment of cartilage repair: a validation study of 2 scoring systems. Arthroscopy 2005;1462–7.
16. Fleiss JL. The Design and Analysis of Clinical Experiments. New York: Wiley, 1986.
17. Bland JM, Altman DG. Cronbach's alpha. BMJ 1997; 572.
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 159–74.
19. Oakley SP, Portek I, Szomor Z, Turnbull A, Murrell GA, Kirkham BW, *et al*. Poor accuracy and interobserver reliability of knee arthroscopy measurements are improved by the use of variable angle elongated probes. Ann Rheum Dis 2002;540–3.
20. Gooding CR, Bartlett W, Bentley G, Skinner JA, Carrington R, Flanagan A. A prospective, randomised study comparing two techniques of autologous chondrocyte implantation for osteochondral defects in the knee: periosteum covered versus type I/III collagen covered. Knee 2006;203–10.
21. Haddo O, Mahroof S, Higgs D, David L, Pringle J, Bayliss M, *et al*. The use of chondrogide membrane in autologous chondrocyte implantation. Knee 2004; 51–5.
22. Brittberg M, Winalski CS. Evaluation of cartilage injuries and repair. J Bone Joint Surg Am 2003;58–69.