TISSUE ENGINEERING
Volume 7, Number 3, 2001
Mary Ann Liebert, Inc.

Validation of a Simple Histological-Histochemical Cartilage Scoring System

SHAWN W. O'DRISCOLL, Ph.D., M.D., F.R.C.S.(C), ROBERT G. MARX, M.D., M.Sc., DORCAS E. BEATON, M.Sc., BScOT., YASUSHI MIURA, M.D., STEPHEN H. GALLAY, M.D., and JAMES S. FITZSIMMONS, B.Sc.

ABSTRACT

In this study, we assessed the validity of a subjective histological-histochemical scoring system as compared to an automated histomorphometry program for analyzing cartilage repair tissue. In the first part of the study, we assessed the ability of the human eye to estimate the percent cartilage in a histological section. Twenty-nine rabbit periosteal explants that had been cultured in agarose transforming growth factor-β (TGF-β) were selected so that the percentage of cartilage in the specimens was distributed equally from 0% to 100%. Color photomicrographs were evaluated by 5 expert observers who gave a visual estimate of the percent cartilage. There was a strong correlation between the estimated and actual percent cartilage ($R^2 = 0.92$, $p < 0.0001$) and among the observers (I.C.C. = 0.89). On average, the estimated percent cartilage was within ten percent of the actual percent measured. In the second part, we compared the data derived using a simple cartilage score with those obtained by automated image analysis. The histological slides from 159 explants cultured under various experimental conditions (14 treatment groups) in two different experiments were analyzed. The cartilage content was estimated visually and a score from 0 to 3 was assigned. A previously validated, computerized image analysis system was used to measure the actual percent cartilage. Statistical analyses revealed a good linear regression ($R^2 = 0.84$, $p = 0.0001$), and even better polynomial correlation between the actual measurement and the score ($R^2 = 0.88$, $p = 0.0001$). These data demonstrate the validity of a simple histological-histochemical subjective scoring system. A computerized automated program such as the one employed in this study is preferable due to its many advantages. However, a subjective scoring system may be appropriate to use when the funding and expertise required for a computerized image analysis program are not available.

The Cartilage and Connective Tissue Research Laboratories, Mayo Clinic, Mayo Foundation, Rochester, Minnesota, and St. Michael's Hospital, University of Toronto, Toronto, Canada.

This paper was presented at the 38th Annual meeting of the Orthopaedic Research Society, February 17–20, 1992, Washington, DC.

INTRODUCTION

With the exponential growth of interest in cartilage repair and tissue engineering, it is becoming increasingly important that we have clear, concise, and precise means of communicating the results of our experimental studies. Histological and histochemical analyses will always play an essential role in assessing neochondrogenesis (new cartilage growth). This can be performed using a subjective histological-histochemical scoring system or by histomorphometry. Scoring systems typically involve the assignment of a categorical grade according to the percentage of the tissue that is estimated visually to represent cartilage. Although histological-histochemical scoring systems have been widely used in orthopedic research to assess the quality or quantity of cartilage,[1–7] they are semiquantitative and subject to inter- and/or intraobserver variability. There are three theoretical criticisms and potential pitfalls of using such a subjective scoring system. The first is that the assessment is subjective based on a visual estimation alone. The second is that the score assigned is a categorical data point (0, 1, 2, or 3) that is a crude estimate. Finally, while it has been assumed that there is a degree of linearity in a relationship between the score assigned and the degree of cartilage formation, this has never been shown. Their application has not been scientifically validated.

Objective methods of assessment include manual histomorphometry and computerized image analysis, which offer greater degrees of objectivity, accuracy, and reproducibility and are therefore more appropriate for statistical analyses.[8] Manual histomorphometry can be performed by tracing, outlining, cutting, and weighing and by linear integration or point-counting. All of these methods are tedious and time-consuming. They are also somewhat observer-dependent. For this reason, interobserver variability can be high as has been found in bone histomorphometry.[9] Computerized histomorphometry has been found to reduce intra- and interobserver variation compared to manual methods for bone histomorphometry.[10,11] Automated histomorphometry, which has found wide acceptance and application in other areas such as engineering, is used commonly for bone histomorphometry.

We have had substantial experience with both histological-histochemical scoring and computerized image analysis for automated histomorphometry in the field of cartilage repair and regeneration.[2,12,13] On the basis of this experience, we designed this experiment to test the following two hypotheses.

Hypothesis

1. The trained human eye is highly capable of estimating the percentage of a histological section that is cartilage.
2. There is good correlation between the data derived using a simple subjective scoring system and those derived using an automated image analysis program.

MATERIALS AND METHODS

Part 1. Visual estimation of percent cartilage

To test the hypothesis that the trained human eye is highly capable of estimating the percentage of a histological section that is cartilage, five researchers (S.W. O'Driscoll, Y. Miura, S.H. Gallay, R.B. Salter, and A.R. Poole), all experienced in cartilage histology, from three different centers were given color photomicrographs of 29 histological sections from rabbit periosteal explants cultured *in vitro* (see below for details of explain method[5–7]). In each case, the instructions were the same. The observer was asked to estimate visually what he thought the percentage of the total area of the section was represented by cartilage. They were not told what criteria to use to define "cartilage" and no reference was made to color or appearance of the matrix, nor of cellular morphology. All assessments were made blind, *i.e.*, without knowledge of the histomorphometry results.

The sections were selected so that the percentage of the total area that was cartilage was equally distributed through the range from 0% to 100%. Additionally, nine of the sections chosen were selected on the basis of the presence of artifacts, mottling or blending of cartilaginous and fibrous tissues, multiple is-

Table 1. Histological-Histochemical Scoring System

Score	Histological-Histochemical Findings Chondrocytes	Safranin O Staining
0	None (or almost none)	None (or almost none)
1	<50%	Slight
2	>50%	Moderate
3	All (or almost all)	Normal (or nearly normal)

lands of cartilage surrounded by noncartilaginous tissue, and/or variable staining with safranin O. The color photomicrographs of these sections, taken in random order with respect to cartilage content, were printed at final magnifications of 100× or 200× so that the whole explant was included in each photograph.

The sections were also analyzed by a different person using a previously published automated computerized image analysis program (see the section, Automated Computerized Histomorphometry below) and the values estimated by each observer compared to the histomorphometric values.[13]

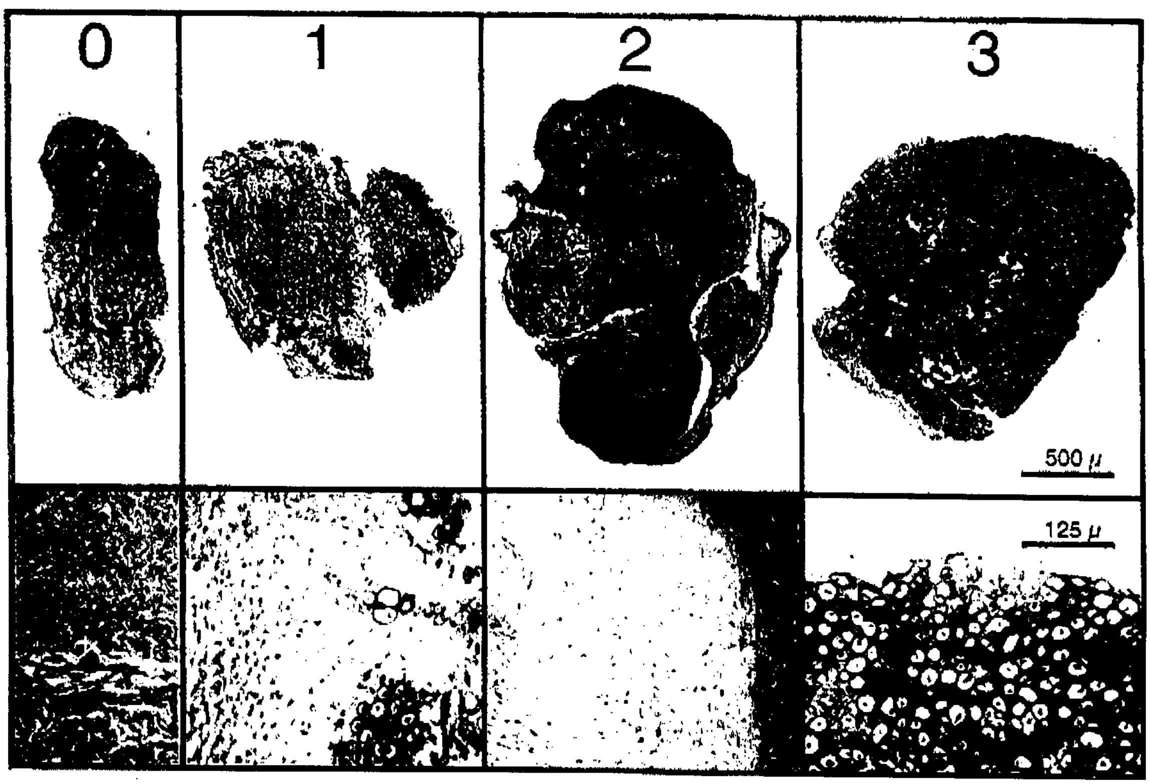

FIG. 1. Simple cartilage score. A score of 0 means no (or almost no) cartilage, whereas a 3 means all (or almost all) cartilage. If the score is not a 0 or a 3, it is either a 1 or a 2, a 1 looking more like a 0 and a 2 looking more like a 3. **(0–3)** Typical examples of specimens that were rated 0, 1, 2, or 3, respectively, unanimously by three observers. Each is a periosteal explant that had been cultured in agarose for 6 weeks and supplemented with 10 ng/ml of TGF-β1 for the first 2 weeks (safranin O/Fast Green stain).

Part 2. Simple cartilage score

To test the hypothesis that there is good correlation between the data derived using a simple subjective cartilage scoring system and those derived using an automated image analysis program, the previously published histological-histochemical scoring system[2] that is based on the cellular morphology and uptake of safranin 0 stain as detailed in Table 1 and illustrated in Fig. 1 was explained to an orthopedic surgeon with no prior experience in cartilage research. It was modified from that originally described by Mankin.[1] He was given only the following instructions. A 0 means no or almost no cartilage, whereas a 3 means all or almost all cartilage. If the score is not a 0 or a 3, it is either a 1 or a 2. To decide which it is, ask yourself, is it more like a 0 or a 3? In the case of the former, the score is 1; in the latter, it is 2.

The histological sections from 159 rabbit periosteal explants cultured in organ culture under various experimental conditions in two different experiments were analyzed. These were used so that not only did we have a spectrum of explants with cartilage ranging from zero to 100%, but also because we could group the explants according the 14 respective experimental groups from which they originally came. This permitted us to compare the mean values for each group obtained using the scoring system with those obtained by the computerized histomorphometry. The reason to do this is that differences in means are more important to experimental investigation than differences in individual values.

To determine agreement among observers using the simple cartilage score, three of the observers independently rated 30 sections and assigned them a score of 0, 1, 2, or 3.

Histological test sections

The model from which we obtained histological sections was an agarose/transforming growth factor-β (TGF-β1) model from the literature that has been shown to promote chondrogenesis in periosteal explants cultured for 6 weeks.[5–7] Sections from the center of each specimen were cut three microns thick and stained with safranin O/fast green.[14]

Automated computerized histomorphometry

This method, developed in our laboratory, has been documented in detail.[13] A Hitachi FP–C1H high-resolution RGB video camera and a VA–C1 RGB adapter were attached to a Zeiss Axioskop microscope and the output sent to a Kontron 486 microcomputer for image digitization, storage, and processing using the Vidas 2.1 Image Analysis software, which was customized for cartilage analysis (all purchased from Carl Zeiss Canada, Don Mills, Ontario M3B 2S6). This software has been designed for real color separation and the camera is high resolution to permit analysis using low-power objectives (1×–2.5×), which is necessary for analyzing large fragments of cartilage. The analysis of cartilage is based on the intensity of safranin O stain. Because this stain is not a pure color, but rather a mixture of primary colors, separate thresholds are set for detecting the red, green, and blue components of the color in the section, and these can be distinguished either automatically or by observer interaction. Threshold adjustment is generally required only when artifacts are erroneously interpreted by the computer, and was not used for the sections analyzed in this study. Holes in the tissue (lacunae or artifacts) are discriminated mathematically on the basis of size and shape (sphericity), calculated separately, and the lacunae included but not the artifacts. This gives a total percentage area of cartilage that is not arbitrarily reduced by the fact that the lacunae do not stain with safranin O. Visual outlines are made on the video image for conformation of the derived data. The lacunae-artifact discrimination is presented visually for observer confirmation. This computerized method has previously been validated for accuracy and precision against a standard method for manual histomorphometry performed by five independent cartilage experts, and has been used extensively with reproducible results.

Statistics

Interobserver variability was assessed for the visual estimates using the intraclass correlation coefficient (ICC).[15] The visual estimates were compared to the values obtained by automated computer histomorphometry using linear regression in part 1 of the study. Linear and polynomial regressions were used to analyze the relationship between the subjective histological-histochemical scores and the measurements made with the image analysis program in part 2 of the study.

RESULTS

Part 1. Visual estimation of percent cartilage

The average percent of the section area that was estimated to be cartilage by each observer ranged from 34 to 45% (Fig. 2). The concordance (agreement) among observers was good, as indicated by the ICC of 0.66. The concordance is considered good if the ICC is 0.4–0.75, and excellent if the ICC > 0.75.[15] We feel that such a range of values (only 10%) would be quite acceptable for measurement error. The average percent cartilage obtained by visual estimation (41%) and that obtained by automated histomorphometry (45%) differed by only 4% (Fig. 3).

There was a strong linear relationship between the visual estimation of the percent cartilage in the 29 specimens and the percent cartilage measured by the computer using the automated histomorphometry program (Fig. 2). For the five observers, the R^2 were 0.65, 0.87, 0.91, 0.92, and 0.93 with the mean $R^2 = 0.92$, and a median $R^2 = 0.91$, respectively. These were highly statistically significant ($p < 0.0001$ for all observers). The variability from one observer to the next related primarily to how they estimated the difficult-to-interpret sections with artifacts and variable staining.

Thus, the eye of an expert is capable of estimating cartilage areas remarkably well. This should not be surprising considering how readily those with substantial experience in mechanics can identify the size of wrench, nut, or bolt, etc., simply by looking at it.

Part 2 Subjective scoring vs. automated image analysis

Linear regression revealed that there was a highly statistically significant relationship between the subjective histological-histochemical score assigned and the percentage of the tissue that was cartilage as de-

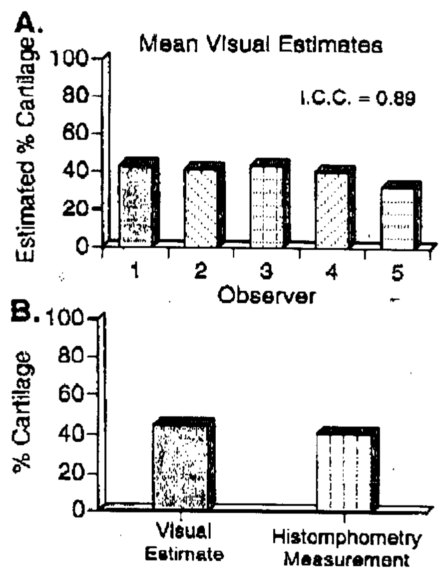

FIG. 2. (A) Five expert observers estimated visually the areas they considered to be cartilage on 29 photomicrographs. This graph represents the means of those values for each observer. There was strong concordance (ICC = 0.89) among them, with a total range of only 11% from the highest to lowest mean values. (B) Comparison of values for 29 sections evaluated by visual estimation by the five observers and by automated computerized histomorphometry.

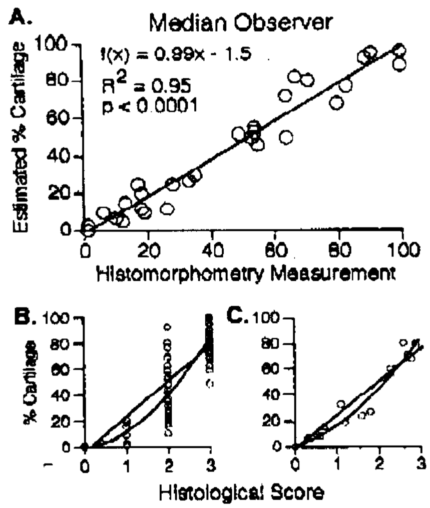

FIG. 3. (A) This graph shows the actual values and the correlation between the visual estimates and histomorphometry measurements for one observer whose correlation represents the median of the five observers. In other words, two of the other observers had better correlations, and two not as good. (B and C) Correlation between simple cartilage scores and histomorphometry measurements for 159 periosteal explants cultured under 14 different experimental conditions in two separate experiments. (B) Raw data. (C) means from 14 different experimental groups were stratified according to group. The correlation was better than that for the raw data. From these data, it is clear that considering comparisons of mean scores across groups would lead to the same ultimate conclusions regardless of method used.

termined by the computerized image analysis program. Simple linear regression yielded the following formula:

Percent cartilage = 27 × score, $R^2 = 0.84$, $p = 0.0001$

The variance in the data was better explained by a second order polynomial regression whose formula was:

Percent cartilage = 8 × score^2 + 4 × score, $R^2 = 0.88$, $p = 0.0001$

The data from fourteen different experimental groups were stratified according to group, and are illustrated in Fig. 3. There was a better correlation between the mean subjective histological score and the mean percentage area of cartilage:

Percent cartilage = 26 × score, $R^2 = 0.94$, $p = 0.0001$

Again this was slightly better with a second order of polynomial regression:

Percent cartilage = 6 × score^2 + 10 × score, $R^2 = 0.96$, $p = 0.001$

As was the case with the individual samples, there was a general trend toward underestimating the percentage area of cartilage and lower subjective scores. From these data, it is clear that considering comparisons of mean scores across groups would lead to the same ultimate conclusions regardless of method used.

When three of the observers were asked to score 30 specimens using the simple cartilage score, all three agreed in 19/30 cases (Table 2). In the remaining 11 cases, at least two observers agreed and the other observer never differed by more than one score/level. In other words, three experts agreed unanimously on

TABLE 2. AGREEMENT AMONG OBSERVERS USING SIMPLE CARTILAGE SCORE

All 3 agreed	*2/3 agreed*	*No agreement*
19	11	0

the score to be assigned 63% of the time and two of three observers agreed on the score the remaining 37% of times.

DISCUSSION

This investigation provides data that validate the use of a simple histological-histochemical subjective scoring system. This would be particularly true for situations requiring comparison of values among groups although there is a surprisingly good quantitative correlation between the histological score assigned and the percentage area of cartilage. The histological-histochemical subjective scoring system is simple and clear, uncomplicated to understand and use and requires no sophisticated or expensive equipment, nor does it require extensive experience to use as indicated in this study. It demonstrates good agreement among observers. Our data have also shown that this can be explained by the fact that the trained human eye is indeed capable of accurate visual estimations of the percent cartilage in a histological specimen. There are two limitations to its use however. The first is that it might not be accurate for any one specimen due to variability in subjective interpretation. The second is that the absolute value of the scores is less reliable than the relative value when comparing amongst groups. Thus, it would not be as valuable as histomorphometry measurements for comparison between studies by different authors. Also, small effects or differences might not be detected by this technique. A computerized automated program such as the one employed in this study eliminates these disadvantages, but, of course, is expensive and requires training and time to use.

From these data, one should be careful not to infer that we are recommending subjective scoring of tissues instead of histomorphometry analyses. Indeed, the advantages of objectively and reproducibility of histomorphometry clearly make it preferable. Our laboratory invested significant time and resources to develop our own automated computerized image analysis system for evaluating cartilage repair tissue at a time when none was commercially available.[13] That system has become our routine method. Nevertheless, scoring systems are still in use, without scientific proof of their validity. Such validation merits documentation.

Although we used explants to evaluate this cartilage score, it was originally developed for evaluating cartilage in articular defects and was used for that purpose in exactly the same manner as described in this study.

In conclusion, there is a strong and statistically significant correlation between the data derived using a simple histological-histochemical subjective scoring system and those derived using a computerized histomorphometry application for image analysis. Although the use of any objective test is always preferable over a subjective one, the results of this experiment validate the use of a subjective scoring system when the financial resources or technical expertise preclude the availability of a computerized image analysis program.

ACKNOWLEDGMENTS

We are grateful to Robin Poole, Ph.D., D.Sc., of the Shriner's Hospital, McGill University, Montreal, Canada, and Robert Salter, O.C., M.D., F.R.S.C., F.R.C.S.(C), of the Hospital For Sick Children, University of Toronto, who along with Shawn W. O'Driscoll, Yasushi Miura, and Stephen H. Gallay scored the photomicrographs in this study. This study was supported in part by research grants to S.W. O'D. from the Medical Research Council of Canada (MA 10879) and Mayo Foundation.

REFERENCES

1. Mankin, H.J., Dorfman, H., Lippiello, L., and Zarins, A. Biochemical and metabolic abnormalities in articular cartilage from osteo-arthritic human hips. II. Correlation of morphology with biochemical and metabolic data. J. Bone Joint Surg. **53-A**, 523, 1971.
2. O'Driscoll, S.W., Keeley, F.W., and Salter, R.B. 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. **70A**, 595, 1988.
3. Shimuzu, T., Videman, T., Shimazaki, K., and Mooney, V. Experimental study on the repair of full thickness articular cartilage defects: effects of varying periods of continuous passive motion, cage activity, and immobilisation. J. Orthoped. Res. **5**, 187, 1987.
4. van der Sluijs, J.A., Geesink, G.T., van der Linden, A.J., Bulstra, S.K., Kuyer, R., and Drukker, J. The reliability of the Mankin score for osteoarthritis. J. Orthoped. Res. **10**, 58, 1992.
5. Miura, Y., Fitzsimmons, J.S., Commisso, C., Gallay, S.H., and O'Driscoll, S.W. Enhancement of periosteal chondrogenesis in vitro: Dose-response for transforming growth factor-beta 1 (TGF-β1). Clin. Orthoped. **301**, 271, 1994.
6. Gallay, S.H., Miura, Y., Commisso, C.N., Fitzsimmons, J.S., and O'Driscoll, S.W. Relationship of donor site to chondrogenic potential of periosteum in vitro. J. Orthoped. Res. **12**, 515, 1994.
7. O'Driscoll, S., Recklies, A., and Poole, A. Chondrogenesis in periosteal explants. An Organ culture model for in vitro study. J. Bone Joint Surg. **76A**, 1042, 1994.
8. Anderson, J.M. Histomorphometry. In: Bancroft, J.D. and Stevens, A., ed. Theory and Practice of Histological Techniques. Edinburgh and New York: Livingstone Churchill, 1982, p. 548.
9. Compston, J.E., Vedi, S., and Stellon, A.J. Inter-observer and intra-observer variation in bone histomorphometry. Calcif. Tissue Int. **38**, 67, 1986.
10. Chavassieux, P.M., Arlot, M.E., and Meunier, P.J. Intermethod variation in bone histomorphometry: comparison between manual and computerized methods applied to iliac bone biopsies. Bone **6**, 221, 1985.
11. Juvin, R., Camus, E., Usson, Y., and Phelip, X. An automated method for bone histomorphometry: assessment with reference to usual static and dynamic parameters. J. Bone Miner. Res. **5**, 235, 1990.
12. O'Driscoll, S.W., Keeley, F.W., and Salter, R.B. 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. **68A**, 1017, 1986.
13. O'Driscoll, S., Marx, R., Fitzsimmons, J., and Beaton, D. A method for automated cartilage histomorphometry. Tissue Eng. **5**, 13, 1999.
14. Rosenberg, L. Chemical basis for the histological rise of safranin O in the study of articular cartilage. J. Bone Joint Surg. **53A**, 69, 1971.
15. Fleiss, J.L. The Design and Analysis of Clinical Experiments. Toronto, John Wiley and Sons: 1986.

Address reprint requests to:
Shawn W. O'Driscoll, Ph.D., M.D., F.R.C.S.(C)
Director, Cartilage and Connective Tissue Research Laboratory
Mayo Clinic, Mayo Foundation
200 1st Street SW, MBS 3-87
Rochester, MN 55905

E-mail: odriscoll.shawn@mayo.edu