Poor accuracy and interobserver reliability of knee arthroscopy measurements are improved by the use of variable angle elongated probes

S P Oakley, I Portek, Z Szomor, A Turnbull, G A C Murrell, B W Kirkham, M N Lassere

Objectives: (a) To determine the accuracy and reliability of arthroscopic measurements of cartilage lesion diameter in an artificial right knee model; (b) to determine whether the use of a set of variable angle elongated probes improves performance; and (c) to identify other sources of variability.

Methods: Ovoid “lesions” were drawn on the five cartilage surfaces of four plastic knees models. Two observers assessed these 20 lesions arthroscopically, measuring two diameters in orientations parallel and orthogonal to the probe. Observer 1 (orthopaedic surgeon) and observer 2 (arthroscopic rheumatologist) made two sets of measurements, firstly with the conventional probe and five months later with the variable angle elongated (VAE) probes. The knees were disarticulated to determine true lesion diameter.

Results: Observer 1 had negligible bias and good accuracy regardless of orientation or probe type. Observer 2 demonstrated both bias and poor accuracy using the conventional probe. Both improved using VAE probes. Poor interobserver reliability with conventional probes also improved using VAE probes. Major sources of variability could be traced to the probe type, the characteristics of the operator, and the orientation of the lesion in relation to the probe; the lesion location itself did not cause variability.

Conclusions: Variation in accuracy and poor interobserver reliability of measurements with conventional methods of cartilage lesion diameter measurement improved when specially designed measurement probes were used. Arthroscopic measurements performed as well as most clinical and radiographic measures. These findings have important implications for the use of arthroscopy as an outcome in multicentre trials where arthroscopists have different levels of experience.

Arthroscopy is a commonly performed procedure that permits direct visualisation and assessment of articular cartilage in the knee. Many arthroscopic methods have been developed to grade the severity of articular cartilage damage. At times arthroscopy has been used to measure outcome in therapeutic trials.1,2 Some grading methods estimate lesion size as a percentage of the articular cartilage area occupied.3 Others require estimation of lesion diameter.4,5 Less is known about the accuracy and reliability of the latter.

Our three objectives were to determine the accuracy and reliability of arthroscopic measurements of cartilage lesion diameter in an artificial knee model, to determine whether the use of a set of variable angle elongated (VAE) probes improves performance, and to identify other sources of variability (lesion location, orientation of measurement, and factors attributable to the operator).

Abbreviations: ANOVA, analysis of variance; ICC, intraclass correlation coefficients; SDD, smallest detectable difference; VAE, variable angle elongated (probes)
lesions seen in actual arthroscopic procedures. The knees were
filled with water and observers performed arthroscopy to
measure the diameter of lesions in orientations parallel and
then orthogonal to the shaft of the probe. Two sets of
measurements were made by each observer with the conven-
tional probe without knowledge of a future study using
improved probes. The knees were then disarticulated and true
lesion diameter determined.

Five months later the same two observers were asked to
repeat the measurements of the same 20 ovoid lesions, now
using VAE probes. The observers were allowed to choose from
among probes with foot processes set at a range of different
angles (0, 30, 45, 60, and 90 degrees). Neither observer saw the
plastic knees in a state of disarticulation. No specific feedback
about the results of measurements with the conventional
probe was given in the interim.

Statistical methods
Three complementary statistical methods were used to evalu-
ate the accuracy of the observed diameter and the reliability of
the measurements within and between observers.

Method 1
Intraclass correlation coefficients (ICCs) were determined
from analysis of the components of variance (ANOVA). Al-
though they are usually used as a measure of reliability, they
can be used as a measure of accuracy when the true lesion
diameter is known. We have used a random effects ICC as
described by Shrout and Fleiss.12 ICCs are a relative measure of
agreement and therefore are biased towards higher coeffi-
cients (1.0 is perfect agreement, 0.0 is no agreement) if the
lesions measured vary over a larger range of values.

Method 2
The 95% confidence limits of the standard deviation of the
difference of the true diameter minus the observed diameter
was also used to estimate accuracy and reliability. The statistic
has been called the smallest detectable difference (SDD)13
and is derived from the limits of agreement method14 and quanti-
fies random error, whereas the mean of the difference scores
(also known as the paired t test) quantifies systematic error
(also known as bias). The ICC cannot discriminate between
random and systematic error. However, the SDD, unlike the
ICC, is an absolute measure of agreement, therefore it is biased
toward smaller values (SDD of 0 is perfect agreement, and
there is no convention that anchors the upper limit) if the
lesions are measured over a narrower range of values.

Accuracy was defined as the observed minus the true
diameter. Thus positive difference scores represent overesti-
mation and negative scores underestimation.

Method 3
ANOVA15 was used to determine which components of the
variance were statistically significant.

RESULTS
With the conventional probe, observer 1, an experienced
orthopaedic surgeon, had overall satisfactory accuracy and
intraobserver reliability (table 1). His ICCs were all greater
than 0.75. He had essentially no bias (three of four mean dif-
ference scores were 0, but there was moderate random error
(SDD ±3 and ±5 for the two orientations) for both accuracy
and reliability.

To illustrate the difference between the ICC and SDD
approaches, it should be noted, firstly, that there was a larger
range of values for the orthogonal (4–36 mm) than the para-
lel (5–17 mm) orientation. The ICC was better (that is, larger)
in the orthogonal than the parallel (0.95 v 0.77), whereas the
SDD was better (that is, smaller) in the parallel than in the
orthogonal (±3 v ±5) orientation.

Observer 2, in contrast, had poor accuracy but satisfactory
intraobserver reliability; in other words he was consistent in
his inaccuracy with the conventional probes (table 1). The
poor accuracy was due to both considerable systematic bias
(underestimation of 4 and 6 mm in the two orientations) and
random error (±5 and ±6). The interobserver reliability was
poor. There were substantial systematic bias and random dif-
fferences between observer 1 and observer 2.

Upon re-test several months later, observer 1 showed mod-
est additional improvement in both accuracy and reliability
with use of VAE probes. Random error, although less, was still
present; its relative importance is discussed later. With this
repeat study, observer 2 showed a substantial improvement in
Table 1  Accuracy and reliability of diameter measurements

<table>
<thead>
<tr>
<th>Observer</th>
<th>Conventional probe</th>
<th>VAE probes</th>
<th>Conventional probe</th>
<th>VAE probes</th>
<th>Conventional probe</th>
<th>VAE probes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC</td>
<td>ICC</td>
<td>ICC</td>
<td>ICC</td>
<td>ICC</td>
<td>ICC</td>
</tr>
<tr>
<td></td>
<td>% SDD (± SDD)</td>
<td>% SDD (± SDD)</td>
<td>% SDD (± SDD)</td>
<td>% SDD (± SDD)</td>
<td>% SDD (± SDD)</td>
<td>% SDD (± SDD)</td>
</tr>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parallel</td>
<td>0.82 (0.82–0.84)</td>
<td>0.93 (0.93–0.99)</td>
<td>0.90 (0.90–0.92)</td>
<td>0.92 (0.92–0.94)</td>
<td>0.96 (0.96–0.98)</td>
<td>0.94 (0.94–0.97)</td>
</tr>
<tr>
<td>Orthogonal</td>
<td>1.4 (1.4–1.5)</td>
<td>0.77 (0.77–0.80)</td>
<td>1.4 (1.4–1.5)</td>
<td>0.77 (0.77–0.80)</td>
<td>1.4 (1.4–1.5)</td>
<td>0.77 (0.77–0.80)</td>
</tr>
<tr>
<td>ICC Diff (mean)</td>
<td>2.12 (1.1)</td>
<td>2.12 (1.1)</td>
<td>2.12 (1.1)</td>
<td>2.12 (1.1)</td>
<td>2.12 (1.1)</td>
<td>2.12 (1.1)</td>
</tr>
<tr>
<td>% SDD (± SDD)</td>
<td>4.11 (0.11–0.11)</td>
<td>4.11 (0.11–0.11)</td>
<td>4.11 (0.11–0.11)</td>
<td>4.11 (0.11–0.11)</td>
<td>4.11 (0.11–0.11)</td>
<td>4.11 (0.11–0.11)</td>
</tr>
<tr>
<td>Interobserver reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICC Diff (mean)</td>
<td>2.12 (1.1)</td>
<td>2.12 (1.1)</td>
<td>2.12 (1.1)</td>
<td>2.12 (1.1)</td>
<td>2.12 (1.1)</td>
<td>2.12 (1.1)</td>
</tr>
<tr>
<td>% SDD (± SDD)</td>
<td>4.11 (0.11–0.11)</td>
<td>4.11 (0.11–0.11)</td>
<td>4.11 (0.11–0.11)</td>
<td>4.11 (0.11–0.11)</td>
<td>4.11 (0.11–0.11)</td>
<td>4.11 (0.11–0.11)</td>
</tr>
</tbody>
</table>

VAE probe, variable angulated elongated probe (see text); accuracy, observed diameter minus true diameter; intraobserver reliability, agreement between 1st versus 2nd sets of estimates by each observer; interobserver reliability, agreement between repeated measurements; diffmean, mean of the difference scores (observed diameter minus true diameter); SDD, standard deviation of the difference scores; (% SDD), SDD expressed as a percentage of accuracy or reliability over time.

DISCUSSION

This study determined the accuracy and reliability and explored sources of variability of arthroscopic measurements of cartilage lesion diameter in an artificial knee model. Overall, with the use of the VAE probes, accuracy and reliability were both good. The percentage SDD varied from 8 to 24%, signifying that arthroscopic measurements performed as well as most clinical and radiographic measures. Accuracy, as determined by agreement statistics (degree of random error), also improved with the use of the VAE probes, and although this improvement was less dependent on the operator, it was influenced somewhat by the method of statistical analysis employed. Intraobserver reliability was also unaltered with the VAE probes but, importantly, interobserver reliability improved considerably.

We showed that systematic over- or underestimation of lesion diameter (bias) with conventional probes was operator dependent, but bias and, consequently, differences between operators decreased substantially with the use of improved (VAE) probes. Accuracy, as determined by agreement statistics (degree of random error), also improved with the use of the VAE probes, and although this improvement was less dependent on the operator, it was influenced somewhat by the method of statistical analysis employed. Intraobserver reliability was also unaltered with the VAE probes but, importantly, interobserver reliability improved considerably.

This study was conducted in a highly artificial, albeit favourable, situation to study the effects of image distortion, among other factors, on arthroscopic measurements. Studies in plastic knees have a number of advantages. Lesions on “cartilage surfaces” were dark coloured with clearly defined margins. In such circumstances one would expect good accuracy and reliability. Comparison with a “gold standard” measurement upon disarticulation may be performed, and measurements of reliability can incorporate variations in procedural technique such as the ability to reproduce the positioning of the arthroscope and the probe. These studies are difficult to undertake in vivo because conditions change during the procedure as synovium becomes oedematous and obscures vision, and cartilage damage may change between
repeated arthroscopic procedures. By the nature of the experiment, observers were aware of the type of probe used. However, observers were unaware that future study was planned with improved probes, they were always unaware of the true lesion size, and the time between measurements using the conventional and VAE probes was large.

Several source of variability were demonstrated. One factor was training and experience of the operator. The better accuracy of observer 1, an orthopaedic surgeon, may be explained by his greater arthroscopic and open knee surgery operating experience. This would improve spatial awareness within the knee joint, thereby compensating for distortions created by arthroscopy. By contrast, observer 2, was a rheumatologist with experience over 200 knee arthroscopies but no experience with open knee surgery and compensation for distortion.

Another factor was orientation of the lesion in relation to the probe. Overall, both systematic and random accuracy were less in the parallel than in the orthogonal orientation. Although this seems counterintuitive, it is probably due to a greater parallax effect because the conventional probe cannot easily be lain flat on the cartilage surface, especially by a less experienced operator. The design of the VAE probe dealt with this difficulty and improved both the accuracy and interobserver reliability. The improved accuracy with the VAE probes may also be due to its finer calibrations (2 mm) and elongated foot processes, which permit easier measurement. However, the greater improvement still occurred in the parallel orientation, not the orthogonal.

These methods are yet to be tested in vivo, where additional challenges will also be encountered. Different grades of cartilage lesion with poorly defined margins are common, so one would expect reduced accuracy and reliability. However, it may be that in this context the true value of the VAE probes is apparent. Furthermore, other easily implemented techniques may be found to improve accuracy and reliability of arthroscopic measurements.

The implications of these types of studies of the fundamentals of measurement are considerable. Arthroscopic outcome measures will continue to play an important part in test-of-concept studies and for validation of full scale clinical trials of agents directed at cartilage protection. These trials are likely to be multicentre and involve arthroscopists of differing backgrounds and differing levels of experience. Although arthroscopy is an invasive procedure, it remains the most sensitive method for detecting early cartilage changes at a time when cartilage protection may have the greatest benefit. It also remains the only method to directly assay physical and biomechanical properties of cartilage, which may be important outcomes themselves. Thus arthroscopy remains of interest as a highly informative outcome measure in arthritis.

**ACKNOWLEDGEMENTS**

Dr Oakley is supported by a research scholarship from the Arthritis Foundation of Australia. We thank the staff at the operating theatres of St George Hospital for the use of their facilities, the St George Hospital Department of Orthopaedic Surgery for the use of the plastic knee teaching models, and Kempsey Valley Arks for their financial support.

---

**Table 2** Effects upon accuracy and reliability analysed by ANOVA

<table>
<thead>
<tr>
<th>Source</th>
<th>Both Observers</th>
<th>Observer 1</th>
<th>Observer 2</th>
<th>Both Observers</th>
<th>Observer 1</th>
<th>Observer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F  p</td>
<td>F  p</td>
<td>F  p</td>
<td>F  p</td>
<td>F  p</td>
<td>F  p</td>
</tr>
<tr>
<td>Model</td>
<td>23.69 0.00</td>
<td>1.97 0.07</td>
<td>21.94 0.00</td>
<td>20.02 0.00</td>
<td>10.34 0.00</td>
<td>12.03 0.00</td>
</tr>
<tr>
<td>Observer 1</td>
<td>139.15 0.00</td>
<td>N/A</td>
<td>33.89 0.00</td>
<td>N/A</td>
<td>33.89 0.00</td>
<td>N/A</td>
</tr>
<tr>
<td>Location</td>
<td>2.19 0.07</td>
<td>1.31 0.27</td>
<td>2.06 0.09</td>
<td>4.78 0.00</td>
<td>2.15 0.08</td>
<td>3.23 0.01</td>
</tr>
<tr>
<td>Orientation</td>
<td>6.64 0.01</td>
<td>3.40 0.07</td>
<td>27.73 0.00</td>
<td>79.31 0.00</td>
<td>53.04 0.00</td>
<td>30.03 0.00</td>
</tr>
<tr>
<td>Probe type</td>
<td>32.14 0.00</td>
<td>3.19 0.08</td>
<td>95.65 0.00</td>
<td>7.85 0.01</td>
<td>0.39 0.53</td>
<td>29.25 0.00</td>
</tr>
<tr>
<td>R^2</td>
<td>0.37 0.07</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Accuracy, difference between true diameter and observed diameter; reliability, repeated measurements; N/A, not applicable.

---

**REFERENCES**


**Authors’ affiliations**

S P Oakley, I Portek, M N Lassere, Department of Rheumatology, St George Hospital, University of New South Wales, Sydney, Australia

Z Szomor, A Turnbull, GACMurrell, Department of Orthopaedics, St George Hospital

B W Kirkham, Department of Rheumatology, Guy’s and St Thomas’ Hospital Trust, London, UK

Correspondence to: Dr S Oakley, Department of Rheumatology, St George Hospital, Gray St, Kogarah, NSW 2217 Australia; z8425086@student.unsw.edu.au

Accepted 7 January 2002

www.annrheumdis.com
Poor accuracy and interobserver reliability of knee arthroscopy measurements are improved by the use of variable angle elongated probes

S P Oakley, I Portek, Z Szomor, A Turnbull, G A C Murrell, B W Kirkham and M N Lassere

Ann Rheum Dis 2002 61: 540-543
doi: 10.1136/ard.61.6.540

Updated information and services can be found at:
http://ard.bmj.com/content/61/6/540

These include:

References
This article cites 12 articles, 2 of which you can access for free at:
http://ard.bmj.com/content/61/6/540#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections
Surgical diagnostic tests (431)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/