Interobserver and intraobserver variability of glenoid track measurements

Adrian K. Schneider, FMH(Orth), MD\textsuperscript{a,b}, Gregory A. Hoy, MBBS, Dip Anat., FRACS(Orth)\textsuperscript{a,c}, Eugene T. Ek, MBBS(Hons), PhD, FRACS(Orth)\textsuperscript{a,c}, Andrew H. Rotstein, MBBS, FRANZCR\textsuperscript{d}, Julie Tate, BSc(Hons), PgCert(CT)\textsuperscript{d}, David McD Taylor, MBBS, MD, MPH, DRCOG, FACEM\textsuperscript{e}, Matthew C. Evans, MBBS, FRACS(Orth)\textsuperscript{a,c,*}

\textsuperscript{a}Upper Limb Unit, Melbourne Orthopaedic Group, Windsor, Victoria, Australia
\textsuperscript{b}Department of Orthopaedic Surgery and Traumatology, Kantonsspital, St. Gallen, Switzerland
\textsuperscript{c}Department of Surgery, Monash University, Clayton, Victoria, Australia
\textsuperscript{d}Victoria House Medical Imaging, Prahran, Victoria, Australia
\textsuperscript{e}Austin Health Emergency Department, Heidelberg, Victoria, Australia

\textbf{Background:} A method of assessing combined glenoid and humeral bone loss in traumatic shoulder instability with an associated treatment protocol was recently published. The aim of this study was to investigate its reliability and reproducibility.

\textbf{Methods:} Seventy-one patients with unilateral anteroinferior shoulder instability underwent computed tomography scans, from which 3-dimensional images were derived. En face views of both glenoid fossae and with 3 views of the humeral head were provided to 4 assessors to determine interobserver reliability. From these measurements, the shoulder was assigned a treatment classification. Two observers repeated their assessments 1 month later to determine intraobserver reliability. For each measurement, the mean coefficient of variability was calculated.

\textbf{Results:} Assessment of glenoid bone loss showed good interobserver (4 observers agreeing in 90.1\% of cases) and also good intraobserver agreement (94\% and 96\%). There was a poor level of interobserver reliability regarding the on-track or off-track classification (72\%). Intraobserver reliability for this measurement was less variable (90\% and 80\%). There was a poor level of agreement between observers (65\%) regarding treatment classification. The coefficient of variability for the Hill-Sachs lesion measured 19.2\%, indicating a high level of variability for this measurement compared with <4\% for all other measures.

\textbf{Conclusion:} Linear bone loss on the glenoid can be measured reliably and reproducibly; however, evaluation of Hill-Sachs lesions demonstrates a high level of variability, and poor interobserver reliability.

\textbf{Level of evidence:} Basic Science Study; Validation of Classification System

© 2017 Journal of Shoulder and Elbow Surgery Board of Trustees. All rights reserved.

\textbf{Keywords:} Hill-Sachs lesion; glenoid bone loss; traumatic shoulder instability; glenoid track; shoulder dislocation; shoulder

Ethical approval for this study was obtained from The Avenue Hospital Human Research Ethics Committee (project 189).

*Reprint requests: Matthew C. Evans, MBBS, FRACS(Orth), Upper Limb Unit, Melbourne Orthopaedic Group, 33 The Avenue, Windsor, Victoria 3181, Australia.

E-mail address: mevans@mog.com.au (M.C. Evans).

1058-2746/$ - see front matter © 2017 Journal of Shoulder and Elbow Surgery Board of Trustees. All rights reserved.

http://dx.doi.org/10.1016/j.jse.2016.09.058
Bipolar bone loss is now well recognized as a cause for recurrent, anteroinferior shoulder instability. The accepted gold standard for quantification of glenoid bone loss is by 3-dimensional (3D) computed tomography (CT) evaluation. A figure of 25% of linear bone loss is indicative of high rates of failure with soft tissue reconstruction.

Although humeral bone lesions (Hill-Sachs defects) are a recognized factor in recurrent instability, the incidence, size, depth, and location of these lesions varies widely. Quantification of Hill-Sachs defects with regard to those parameters have all been described, but correlation among the various classifications and any treatment algorithm has been lacking. Currently there is no accepted gold standard for quantification of Hill-Sachs lesions and the consequent implications for treatment.

A new treatment paradigm for traumatic anteroinferior shoulder instability has been published by Di Giacomo et al, based on measurements incorporating linear glenoid bone loss and the relationship of the humeral lesion to the “glenoid track”, using CT scans with 3D reconstructions. The aim of this study was to determine the utility of this new treatment paradigm by assessing the interobserver and intraobserver reliability of the measurements on 3D CT scans and the use in guiding the management of patients with anteroinferior instability.

Materials and methods

Type of study

This is a Level I study to test a previously described classification system based on measurements from a radiologic diagnostic test.

Cohort and inclusion criteria

A consecutive series of patients fitting the inclusion criteria was identified. Inclusion criteria were patients with unilateral, anteroinferior glenohumeral instability who had undergone standardized, preoperative CT scanning for assessment of bone loss.

Exclusion criteria

Patients with multidirectional instability, bilateral traumatic instability or previous surgery other than an arthroscopic Bankart repair were excluded.

Data collection methods

We looked at CT scans that were performed at 1 institution (Victoria House Medical Imaging), from January 2012 to October 2015. All patients underwent bilateral CT imaging of both shoulders using an Optima 660 64-slice CT scanner (GE Healthcare, Waukesha, WI, USA). Further technical details of CT imaging are available in the Supplementary Data (Appendix 1).

A GE workstation was used to acquire 3D volume-rendered images of the whole shoulder girdle. Multiplanar reconstructions of the affected shoulder were created in the axial, coronal, and sagittal plane in relation to the glenoid.

After digital subtraction of the humeral heads, 2 en face views of the unaffected and affected glenoids were generated (Fig. 1, a and b). On the pathologic side, 3 images of the disarticulated humeral head were obtained: a posterior view, an oblique-posterior view, and a superior view (12-o’clock) to visualize and quantify any Hill-Sachs lesions (Fig. 2, a, b, and c). All of these reconstructions were produced by an independent and thoroughly trained orthopaedic fellow (A.S.).

Method of interobserver testing

Hard-copy images for each shoulder were provided to each of the 4 assessors, blinded to the identity of the patients. The assessors were 3 shoulder specialist surgeons (M.E., G.H., E.E.) and one musculoskeletal radiologist (A.R.).

Oral and written instructions were provided to all assessors, with a standardized measurement protocol to follow, to minimize variability in measurement technique. All assessors discussed and agreed with the prescribed measurement technique. All assessors performed the assessment independently, without discussion or communication during assessment. Details of the protocol are provided in the Supplementary Data (Appendix 2).

The treatment algorithm of Di Giacomo et al states that all patients with less than 25% linear glenoid bone loss and an on-track Hill-Sachs lesion would receive an arthroscopic stabilization procedure (group 1), and for those with an off-track Hill-Sachs lesion, remplissage would be added (group 2). Patients with 25% linear glenoid bone loss or greater and an on-track Hill-Sachs lesion would receive a Latarjet procedure (group 3), and for those with an off-track Hill-Sachs lesion, remplissage or humeral bone grafting would be added (group 4).

Method of intraobserver testing

Two assessors (M.E., A.R.) repeated the measurement protocol for all shoulders 1 month later, blinded to their previous results, to assess for intraobserver reliability across the same measurements and classification parameters. The order of the patients was randomized for each evaluation to avoid bias.

All findings were tabulated for statistical analysis. An independent statistician (D.T.) analyzed the data. No sample size calculations were undertaken because the study was descriptive, involved no statistical comparison of groups, and enrolled all available patients. The effect sizes in this initial study (eg, proportions) will help to inform sample size calculations in subsequent studies.

Results

Demographics

The CT scans of 71 patients with unilateral traumatic anteroinferior shoulder instability were included. Demographic data, including age, sex, and side were collected. Patients who had not undergone previous arthroscopic stabilization surgery for their shoulder instability were classified...
as “primary cases,” and those who had were classified as “re-
vision cases.” These data are summarized in Table I.

**Interobserver reliability**

Interobserver levels of agreement about whether there was
greater than or less than 25% linear, glenoid bone loss:

- 4 observers agreed in 64 cases (90.1%)
- 3 observers agreed in 5 cases (7.1%) and 1 disagreed
- 2 observers agreed in 2 cases (2.8%) and 2 disagreed

These results indicated there was a good level of agree-
ment between the 4 observers regarding linear, glenoid bone
loss.

Interobserver levels of agreement about whether the com-
bined bone lesion was “on-track” or “off-track”:

- 4 observers agreed in 51 cases (71.8%)
- 3 observers agreed in 14 cases (19.7%) and 1 disagreed
- 2 observers agreed in 6 cases (8.5%) and 2 disagreed

These results indicated there was a poor level of agree-
ment among different observers regarding the classification
of the combined bone loss.

**Intraobserver reliability**

Intraobserver agreement about whether there was greater than
or less than 25% linear, glenoid bone loss:

The first observer (M.E.) agreed with his first conclusion
in 68 of cases (95.8%)

Figure 1  (a and c) Images demonstrate the glenoid on the unaffected side. Glenoid track of the normal glenoid (GT1) can be calculated by multiplying the anteroposterior width (CDg) by 0.83. (b and d) Images show the affected glenoid, with (d) minor erosive bone loss. Glenoid track of the affected side (GT2) is reduced by the amount of bone loss (GT1-d).
The second observer (A.R.) agreed with his first conclusion in 67 of cases (94.4%).

This indicated there was good correlation within each observer on the repeat evaluation regarding linear glenoid bone loss.

Intraobserver agreement about whether the case was on or off track:

The first observer (M.E.) agreed with his own initial analysis in 57 of cases (80.3%)
The second observer (A.R.) agreed with his own initial analysis in 64 of cases (90.1%)

There is less variability for repeated evaluation by the same observer than between observers when assessing the classification of combined bone loss.

Figure 2  Three different positions of the humerus were constructed: (a and d) posterior view, (b and e) oblique-posterior view, and (c and f) superior view. Observers then had to outline the medial edge of the cuff insertion (yellow), the border of the Hill-Sachs lesion (red shaded), and the length of Hill-Sachs interval (HSI, red), on the posterior view. The glenoid track (GT2) of the affected glenoid was then superimposed on the humeral head. Finally, a statement had to be made whether the Hill-Sachs lesion was on-track or off-track.
Our results demonstrate a good level of reliability between and within observers for this measurement.

Discussion

This study tested the reliability and reproducibility of standardized 3D CT measurements to derive an algorithm of treatment for traumatic, anteroinferior shoulder instability. 3D CT is now considered the gold standard diagnostic test measuring linear glenoid bone loss through fracture or bone erosion.\(^1\) Our results demonstrate a good level of reliability between and within observers for this measurement.

When individually assessed, all glenoid-related measurements demonstrated a low CoV, but the measurement of the Hill-Sachs interval showed a high level of variability. This accounts for the poor correlations both between and within observers for the assessment of “on-track” vs. “off-track” lesions, as well as subsequent classification according to the Di Giacomo treatment algorithm.\(^2\)

Our literature review as part of the background for this study revealed several publications with varying methods for describing and quantifying Hill-Sachs lesions.\(^2,16,17\) To date, no single method has been universally accepted.\(^1,7,10,12-15,23,25-29\)

The “glenoid track” theory\(^3\) and the more recently published treatment algorithm\(^4\) that now accompanies it combines the well-accepted method of measuring linear glenoid bone loss and then deriving the width of the “glenoid track.” This measurement is then superimposed on a 3D CT reconstruction of the posterior surface of the proximal humerus along with markings to outline the medial attachment of the rotator cuff and the margins of the Hill-Sachs defect.

Several steps are required for the humeral part of this evaluation, and in our opinion, each one carries potential for error, and each error can compound, leading to the poor reliability and reproducibility that we have identified.

We can find no standardized description of the CT image that should be created. When a 3D rendering of a structure is reproduced on a 2D surface (the computer screen), any small alteration in orientation can lead to a change in the perspective and dimensions of that structure. Furthermore, drawing a line in 2 dimensions to measure a 3 dimensional distance is problematic for the same reasons. In designing our study, we could find no guidelines on the achieving the exact orientation for each scan, and so we endeavored to provide 3 images of varying orientations for the observers to aid in their evaluation of the bone lesion. We also limited our study to a single scanner, with the same software package for all scans, to avoid any variability that such changes might introduce.

Despite these attempts at standardization of scanner and images, there was still marked variation in the determination of the site of cuff insertion and in the margins of the Hill-Sachs lesion. All observers commented to the lead investigator after the analysis regarding difficulties with identifying both of these parameters on the humeral images. Although with repeat observations the individual observers achieved better consistency with their own observations, this does not necessarily mean that they were correct, just consistent!

With 3D imaging, the Hill-Sachs lesion conspicuity is proportional to the depth of the lesion. Therefore, with shallow Hill-Sachs lesions, the medial margin of the Hill-Sachs lesion was more difficult to define. The main challenge with the CT Hill-Sachs interval measurement is delineating the medial margin of the rotator cuff insertion. Unfortunately the soft

\(\text{Table I} \quad \text{Demographic data}
\)

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Primary cases</th>
<th>Revision cases</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male, No. (%)</td>
<td>n = 46</td>
<td>n = 25</td>
<td>N = 71</td>
</tr>
<tr>
<td>Female, No. (%)</td>
<td>39 (85)</td>
<td>21 (84)</td>
<td>60 (84)</td>
</tr>
<tr>
<td>Right side, No. (%)</td>
<td>7 (15)</td>
<td>4 (16)</td>
<td>11 (16)</td>
</tr>
<tr>
<td>Left side, No. (%)</td>
<td>27 (59)</td>
<td>14 (56)</td>
<td>41 (58)</td>
</tr>
<tr>
<td>Age, mean (y)</td>
<td>25.4</td>
<td>25.8</td>
<td>25.5</td>
</tr>
</tbody>
</table>

\(\text{Table II} \quad \text{Coefficient of variability}
\)

| Measurement | CoV (Coefficient of variability: A1B1, height of the unaffected glenoid; A2B2, height of the affected glenoid; CDg, width of the unaffected glenoid; CDb, width of the affected glenoid.]
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A1B1</td>
<td>4.0</td>
</tr>
<tr>
<td>A2B2</td>
<td>3.7</td>
</tr>
<tr>
<td>CDg</td>
<td>2.7</td>
</tr>
<tr>
<td>CDb</td>
<td>3.6</td>
</tr>
<tr>
<td>Hill-Sachs</td>
<td>19.2</td>
</tr>
</tbody>
</table>

Intraobserver agreement with regard to the selection of appropriate surgical treatment based on the classification of Di Giacomo et al:

The first observer (M.E.) agreed with his initial treatment decision in 54 of cases (76.1%) and the second observer (A.R.) agreed with his initial treatment decision in 61 of cases (85.9%).

Again, although there was less variability for the repeated evaluation by the same observer than between observers, the level of agreement overall was poor.

Coefficients of variability

The CoVs were calculated to determine how variable each measurement was around its mean value. The CoV is defined as the ratio of the standard deviation (SD) to the absolute value of the mean, expressed as a percentage:

\(\text{CoV} = 100 \times \frac{\text{SD}}{\text{Mean}}\)

Table II demonstrates the CoV for each of the measurements taken from the 3D CT images. Quantification of the Hill-Sachs interval (Fig. 2, d) demonstrated a 5-fold higher CoV than any other measurement.

\(\text{Discussion}
\)

This study tested the reliability and reproducibility of standardized 3D CT measurements to derive an algorithm of treatment for traumatic, anteroinferior shoulder instability. 3D CT is now considered the gold standard diagnostic test measuring linear glenoid bone loss through fracture or bone erosion. Our results demonstrate a good level of reliability between and within observers for this measurement.

When individually assessed, all glenoid-related measurements demonstrated a low CoV, but the measurement of the Hill-Sachs interval showed a high level of variability. This accounts for the poor correlations both between and within observers for the assessment of “on-track” vs. “off-track” lesions, as well as subsequent classification according to the Di Giacomo treatment algorithm.\(^2\)

Our literature review as part of the background for this study revealed several publications with varying methods for describing and quantifying Hill-Sachs lesions.\(^2,16,17\) To date, no single method has been universally accepted.\(^1,7,10,12-15,23,25-29\)

The “glenoid track” theory\(^3\) and the more recently published treatment algorithm\(^4\) that now accompanies it combines the well-accepted method of measuring linear glenoid bone loss and then deriving the width of the “glenoid track.” This measurement is then superimposed on a 3D CT reconstruction of the posterior surface of the proximal humerus along with markings to outline the medial attachment of the rotator cuff and the margins of the Hill-Sachs defect.

Several steps are required for the humeral part of this evaluation, and in our opinion, each one carries potential for error, and each error can compound, leading to the poor reliability and reproducibility that we have identified.

We can find no standardized description of the CT image that should be created. When a 3D rendering of a structure is reproduced on a 2D surface (the computer screen), any small alteration in orientation can lead to a change in the perspective and dimensions of that structure. Furthermore, drawing a line in 2 dimensions to measure a 3 dimensional distance is problematic for the same reasons. In designing our study, we could find no guidelines on the achieving the exact orientation for each scan, and so we endeavored to provide 3 images of varying orientations for the observers to aid in their evaluation of the bone lesion. We also limited our study to a single scanner, with the same software package for all scans, to avoid any variability that such changes might introduce.

Despite these attempts at standardization of scanner and images, there was still marked variation in the determination of the site of cuff insertion and in the margins of the Hill-Sachs lesion. All observers commented to the lead investigator after the analysis regarding difficulties with identifying both of these parameters on the humeral images. Although with repeat observations the individual observers achieved better consistency with their own observations, this does not necessarily mean that they were correct, just consistent!

With 3D imaging, the Hill-Sachs lesion conspicuity is proportional to the depth of the lesion. Therefore, with shallow Hill-Sachs lesions, the medial margin of the Hill-Sachs lesion was more difficult to define. The main challenge with the CT Hill-Sachs interval measurement is delineating the medial margin of the rotator cuff insertion. Unfortunately the soft
Conclusion

Our study demonstrates that shoulder surgeons and radiologists can reliably and reproducibly measure linear bone loss on the glenoid through the use of standardized 3D CT images. However, with regard to proximal humeral assessment, use of the medial edge of the cuff insertion and the margins of the Hill-Sachs defect introduces a high level of variability in assessment between and within observers. This subsequently results in poor reliability of the “glenoid-track” measurement and the related treatment algorithm. Because of this we cannot recommend the “glenoid-track” method be used to infer treatment for glenohumeral instability.

We conclude that further study is required to develop an accurate and reproducible method of identifying the margins of the Hill-Sachs lesion on 3D CT, and its relationship to the glenoid-track.

Disclaimer

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

Appendices

Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jsse.2016.09.058.

References


