Comparative Study of Four Maxillofacial Trauma Scoring Systems and Expert Score

Chen Chen, DDS,* Yi Zhang, PhD, MD, DDS,† Jin-gang An, DDS, MD,‡ Yang He, DDS, MD,§ and Xi Gong, DDS‖

Purpose: To select a scoring system suitable for the scoring of maxillofacial trauma by comparing 4 commonly used scoring systems according to expert scoring.

Patients and Methods: Twenty-eight subjects who had experienced maxillofacial trauma constituted the study cohort. Four commonly used systems were selected: New Injury Severity Score (NISS), Facial Injury Severity Scale (FISS), Maxillofacial Injury Severity Score (MFISS), and Maxillofacial Injury Severity Score (MISS). Each patient was graded using these 4 systems. From the experience of our trauma center, an expert scoring table was created. After the purpose and scheme of the study had been explained, 35 experts in maxillofacial surgery were invited to grade the injury of the 28 patients using the expert scoring table according to their clinical experience. The results of the 4 scoring systems and expert score were compared.

Results: The results of the 4 scoring systems and expert score demonstrated a normal distribution. All results demonstrated significant differences (P < .01). The Pearson correlation coefficient between the MFISS and expert score was the greatest (0.801). The correlation coefficient between the NISS, FISS, and MISS and the expert score was 0.714, 0.699, and 0.729, respectively. Agreement between the standardized scores and the expert score was evaluated using Bland-Altman plots; the agreement between the standardized MFISS and expert score was the best.

Conclusions: Compared with the other 3 scoring systems, the correlation and agreement between the MFISS and expert score was greater. This finding suggests that the MFISS is more suitable for scoring maxillofacial injuries.

© 2014 American Association of Oral and Maxillofacial Surgeons

Scoring of injury severity is an important part of trauma research. It is an effective method of evaluating the prognosis of patients who have experienced trauma. The Abbreviated Injury Scale (AIS),¹ based on the anatomy, was first proposed in 1971 and has been revised repeatedly. In 1974, Baker et al² found that the severity and mortality of the injury changed regularly with the sum of the square of the 3 greatest AIS grades in 3 different body areas (this rule still holds true for multiple injuries), and the Injury Severity Score (ISS) was proposed. The AIS-ISS system has become the most widely used scoring system in the world. The main indicator of the AIS-ISS scoring system is the probability of survival. The direct threat of lethality from maxillofacial trauma is low. However, injuries can damage the appearance and function of patients and lead to permanent disability and psychological harm.³ Therefore, the AIS-ISS scoring system is not suitable for the assessment of the severity of maxillofacial trauma. Specialists in maxillofacial surgery have established various injury scoring systems according to the characteristics of maxillofacial trauma; however, none has...
been as widely accepted and used as the AIS-ISS scoring system. The main reason has been no reference standard is available for grading systems of injury severity. In maxillofacial trauma, all the complications are focused on dysfunction and facial deformities (similar to the situation with hand injuries). In 1996, 5 hand injury specimens were selected and graded by 25 hand surgeons. Campbell and Kay modified the Hand ISS in line with the reference standard set by the 25 experienced surgeons. Catapano et al performed a similar study in 2010 using a Facial Fracture Severity Scale (FFSS). Hence, expert opinions are very important in the assessment of injury severity.

In the present study, 4 commonly used systems were selected to grade injury severity: the New Injury Severity Score (NISS), Facial Injury Severity Scale (FISS), Maxillofacial Injury Severity Score (MFISS), and Maxillofacial Injury Severity Score (MISS). We then graded the injury severity of 28 patients using these systems. Experts in maxillofacial surgery also graded the injury severity of these patients. The results were analyzed to identify the scoring system that had results most consistent with the expert score.

### Patients and Methods

#### Patients

From February to September 2013, we selected 28 inpatients with maxillofacial trauma (25 males and 3 females; 3 to 64 years old) from the Peking University School and Hospital of Stomatology (Beijing, China). Our institutional ethics committee approved the study, and all patients provided written informed consent to participate. The inclusion criteria were a definitive diagnosis of maxillofacial trauma and detailed description of the physical examination; preoperative imaging data and facial and occlusion photographs available; at preregistration, the soft tissue injury had occurred less than 24 hours previously and the maxillofacial fracture less than 3 weeks previously. All 28 inpatients underwent surgery by the same 2 surgeons of our research team. The injury types of the 28 patients are listed in Table 1.

#### METHODS

**Scoring Method Used by Experts in Maxillofacial Surgery**

We designed an expert scoring table. The items included injury site, injury type, surgical procedure complexity, and predicted complications. Each item was graded from 0 to 5 (with 5 the most severe). The information relating to our 28 patients was sent to 35 experts in maxillofacial surgery by electronic mail (email). All 35 experts were professors or associate professors of maxillofacial surgery in public hospitals in China. They had more than 5 years of clinical experience in maxillofacial trauma and had undertaken more than 100 surgical procedures on trauma patients annually. The experts provided scores for the 4 parameters according to their clinical experience to provide a final score for the expert scoring table. The completed table was then returned to us.

For example, a 31-year-old female experienced a right condylar fracture after falling. We emailed the information related to the patient’s specialized physical examination, preoperative facial, mouth-opening, and occlusion photographs (Fig 1), and imaging data (Fig 2) to the expert. The expert scoring table was also provided. Next, the expert rated the patient by referring to the patient information and graded the injury severity in the expert scoring table according to their experience. Finally, the completed tables were returned to us by email (Table 2). The results for each patient from all the experts were then calculated.

**Grading of Injury Severity Using 4 Scoring Systems**

Three experts in maxillofacial trauma from our research team graded the 28 patients using the NISS, FISS, MFISS, and MISS. The AIS standard used in all the scoring systems was the 2005 version. The mean value of the scores from the 3 experts was considered the final score of each patient.

#### STATISTICAL ANALYSIS

The data were analyzed using the Statistical Package for Social Sciences, version 11.0 (SPSS, Chicago, IL). Using 95% confidence intervals, the 2 largest deviations in the score of each patient provided by the 35 experts were removed. The mean value for each case provided by the remaining 33 experts was then defined as the final score.

### Table 1. INJURY TYPE FOR STUDY COHORT OF 28 PATIENTS

<table>
<thead>
<tr>
<th>Injury Type</th>
<th>Patients (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soft tissue</td>
<td>2</td>
</tr>
<tr>
<td>Mandible fracture</td>
<td>12</td>
</tr>
<tr>
<td>Maxillary fracture</td>
<td>2</td>
</tr>
<tr>
<td>Maxillary and mandible fracture</td>
<td>3</td>
</tr>
<tr>
<td>Fracture of zygoma and zygomatic arch</td>
<td>2</td>
</tr>
<tr>
<td>Fracture of zygoma, zygomatic arch and maxilla</td>
<td>4</td>
</tr>
<tr>
<td>Fracture of zygoma, zygomatic arch and mandible</td>
<td>1</td>
</tr>
<tr>
<td>Panfacial fracture</td>
<td>2</td>
</tr>
</tbody>
</table>

The results of the expert score and those using the 4 scoring systems were tested to determine whether they had a normal distribution. We calculated the Pearson correlation coefficients among the expert score, 4 scoring methods, cost of the operation, and operation time.

The results of 4 scoring systems were standardized according to the range of expert scores (if a case was graded “A” using 1 scoring system but the maximum of the scoring system was “B,” the standardized score of that case according to the expert score was calculated using the equation: $X = A \times \frac{20}{B}$.) The paired-sample $t$ test was used for the statistical analyses. Bland-Altman plots were used to assess the agreement between the scores derived from the 4 standardized scoring methods and those by the experts.

Results

The results of grading the 28 patients by the 35 experts and the 4 scoring systems are listed in Tables 3 and 4. All data had a normal distribution. The correlation coefficients among the expert score, 4 scoring systems, cost of operation, and operation time suggested that the expert score had the greatest correlation coefficients compared with the other 4 scoring systems (Table 5). Examination of the correlation coefficients between the 4 scoring systems and the expert score showed that the Pearson correlation coefficient between the MFISS and expert score was the greatest (0.801). The correlation coefficient between the NISS, FISS, and MISS and the expert score was 0.714, 0.699, and 0.729, respectively (Table 6). All results were statistically significant ($P < .01$).

The results of scoring using the 4 scoring systems were standardized according to the range of the expert score. Agreement between the standardized scores and the expert score was evaluated using Bland-Altman plots (Tables 7, 8; Figs 3-10). The paired sample $t$ test was applied between the “difference” and “ratio” results from the Bland-Altman plots. These data suggested that no significant differences were
present between the standardized NISS (SNISS) and standardized MFISS (SMFISS) nor between the standardized FISS (SFISS) and standardized MISS (SMISS), with regard to the difference and ratio \((P > .05)\). However, the remaining pairwise comparisons had significant differences \((P < .01)\). Next, we computed the data using NCSS-PASS, version 11.0 (NCSS Statistical Software, Kaysville UT). The results showed that the power was greater than 0.9, with a significance level of 0.05 \((\alpha = 0.05)\) when significant differences were present between SNISS and SMFISS and between SFISS and SMISS. The existence of a significant difference between SNISS and SMFISS, or between SFISS and SMISS, was not observed in the present study and must be confirmed by additional research. The agreement between the standardized MFISS and expert score was the closest.

**Discussion**

The AIS-ISS has formed the basis of the trauma scoring systems. The AIS can be used to evaluate trauma severity but not to predict the outcome of injury, because it was based on the anatomic description of the injury.\(^1\)
The ISS is defined as the sum of the squares of the 3 greatest AIS values of 3 body regions. If multiple injuries are present in the same region of the body, only 1 of the AIS values is used, and the survival rate is used as the index. The indexes of several other scoring systems are also associated with the survival rate.

The ISS cannot be used to distinguish single and multiple trauma if assessing injury, and Osler et al described a new method, the NISS, in 1997.

<table>
<thead>
<tr>
<th>Table 2. EXPERT SCORE TABLE</th>
<th>Example of Expert’s Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>Patient number 1</td>
</tr>
<tr>
<td></td>
<td>Injury site</td>
</tr>
<tr>
<td></td>
<td>Injury type</td>
</tr>
<tr>
<td></td>
<td>Complexity of surgical procedure</td>
</tr>
<tr>
<td></td>
<td>Predicted complications</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>Each item was graded from 0 to 5, with 5 the most severe. Chen et al. Four Maxillofacial Trauma Scoring Systems. J Oral Maxillofac Surg 2014.</td>
<td></td>
</tr>
</tbody>
</table>

| Table 3. GRADING OF INJURY SEVERITY OF 28 PATIENTS BY EXPERTS |
|---------------------|---------------------|
| Pt. No.             | Expert Score        |
| 1                   | 6.70                |
| 2                   | 7.30                |
| 3                   | 8.15                |
| 4                   | 6.85                |
| 5                   | 5.30                |
| 6                   | 6.73                |
| 7                   | 12.15               |
| 8                   | 4.21                |
| 9                   | 8.03                |
| 10                  | 10.12               |
| 11                  | 12.00               |
| 12                  | 9.39                |
| 13                  | 4.85                |
| 14                  | 9.03                |
| 15                  | 8.88                |
| 16                  | 11.94               |
| 17                  | 15.09               |
| 18                  | 10.58               |
| 19                  | 9.97                |
| 20                  | 12.61               |
| 21                  | 10.76               |
| 22                  | 9.39                |
| 23                  | 18.30               |
| 24                  | 15.00               |
| 25                  | 13.94               |
| 26                  | 8.79                |
| 27                  | 11.97               |
| 28                  | 8.42                |

<table>
<thead>
<tr>
<th>Table 4. GRADING OF INJURY SEVERITY OF 28 PATIENTS BY THREE EXPERTS IN OUR RESEARCH TEAM USING FOUR SCORING SYSTEMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pt. No.</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>17</td>
</tr>
<tr>
<td>18</td>
</tr>
<tr>
<td>19</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>21</td>
</tr>
<tr>
<td>22</td>
</tr>
<tr>
<td>23</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>26</td>
</tr>
<tr>
<td>27</td>
</tr>
<tr>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5. PEARSON CORRELATION COEFFICIENTS BETWEEN THE FIVE SCORING METHODS AND COST OF OPERATION AND OPERATION TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>---------------------------------</td>
</tr>
<tr>
<td>Cost of operation</td>
</tr>
<tr>
<td>Operation time</td>
</tr>
</tbody>
</table>

Abbreviations: FISS, Facial Injury Severity Scale; MFISS, Maxillofacial Injury Severity Score; MISS, Maxillofacial Injury Severity Score; NISS, New Injury Severity Score; Pt. No., patient number.


The ISS cannot be used to distinguish single and multiple trauma if assessing injury, and Osler et al described a new method, the NISS, in 1997. The NISS
is defined as the sum of the square of the 3 greatest AIS scores in 3 body regions (including those in the same area). Several studies have reported the NISS to be superior to the ISS for the prediction of injury severity and mortality. However, the NISS has not been as widely used or recognized as the ISS.

According to the characteristics of maxillofacial trauma, investigators have established various scoring systems to grade maxillofacial injury. Maxillofacial fracture is the focus of maxillofacial injury scoring systems. The indexes of the AIS-ISS cannot be used to characterize the peculiarities of the severity of maxillofacial injury. Hence, using the characteristics of oral and maxillofacial trauma, experts have developed various scoring methods by giving priority to skeletal injuries, such as the Craniofacial Disruption Score (CDS), FFSS, and ZS scoring system. The CDS divides the craniofacial bone into 20 major regions, with each region divided into several parts. Each anatomic region receives a grade from 0 to 3, which represents the fracture severity. The greatest score for each major region is 5 points. The FFSS documents the fracture patterns in 41 major parts, with each part scoring 0 to 3 points and a maximum score of 123. By summarizing the commonly used scoring methods, the ZS scoring system improved the classification of fractures, increased the weight coefficient of comminuted fractures, and perfected the classification of the teeth, nasal-orbital-ethmoid fractures, and condylar fractures. These scoring methods have constantly been improving the classification of fractures. However, the basis of these systems has been bone classification, and none refer to the facial soft tissue. Furthermore, even the creators of the scoring systems believed that

### Table 6. PEARSON CORRELATION COEFFICIENTS FOR SCORES OF 28 PATIENTS BY EXPERTS AND USING THE FOUR SCORING SYSTEMS

<table>
<thead>
<tr>
<th>Scoring System</th>
<th>Pearson Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>NISS</td>
<td>0.714</td>
</tr>
<tr>
<td>FISS</td>
<td>0.699</td>
</tr>
<tr>
<td>MFISS</td>
<td>0.801</td>
</tr>
<tr>
<td>MISS</td>
<td>0.729</td>
</tr>
</tbody>
</table>

The $P$ value (2-tailed) for the Pearson correlation coefficient in all cases was < .001.

### Table 8. MEAN AND STANDARD DEVIATION OF THE STANDARDIZED FOUR SCORING SYSTEMS AND EXPERT SCORE WITH RESPECT TO “RATIO”

<table>
<thead>
<tr>
<th>Scoring System</th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNISS</td>
<td>0.41 ± 0.15</td>
</tr>
<tr>
<td>SFISS</td>
<td>0.23 ± 0.09</td>
</tr>
<tr>
<td>SMFISS</td>
<td>0.44 ± 0.19</td>
</tr>
<tr>
<td>SMISS</td>
<td>0.25 ± 0.10</td>
</tr>
</tbody>
</table>

The comparison between SNISS and SMFISS and between SFISS and SMISS resulted in $P > .05$; the other pairwise comparisons resulted in $P < .01$. 

### Table 7. MEAN AND STANDARD DEVIATION OF THE STANDARDIZED FOUR SCORING SYSTEMS AND EXPERT SCORE WITH RESPECT TO “DIFFERENCE”

<table>
<thead>
<tr>
<th>Scoring System</th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNISS</td>
<td>−5.86 ± 2.67</td>
</tr>
<tr>
<td>SFISS</td>
<td>−7.68 ± 2.53</td>
</tr>
<tr>
<td>SMFISS</td>
<td>−5.37 ± 1.96</td>
</tr>
<tr>
<td>SMISS</td>
<td>−7.33 ± 2.37</td>
</tr>
</tbody>
</table>

Comparison between SNISS and SMFISS and between SFISS and SMISS resulted in $P > .05$; the other pairwise comparisons resulted in $P < .01$.

### FIGURE 3. Bland-Altman plot showing the difference between the standardized New Injury Severity Score (SNISS) and expert score.

evaluation of soft tissue injuries should also be included.5,8

The FISS9 includes the classification of laceration of the facial soft tissue and that of bone. In FISS, the facial bones have been divided into the upper, middle, and lower thirds, with 1 item for facial laceration added. However, the classification of bones is not sufficiently detailed and cannot be used to distinguish displaced and comminuted fractures. Subsequent scoring systems such as the MFISS3 and MISS,10 in addition to anatomic damage, also classified the impairment of maxillofacial function and facial appearance, which can reflect the effect on patient quality of life (QoL) caused by maxillofacial injuries. These 2 scoring systems were based on the AIS and combine the ISS parameters of maxillofacial function and appearance (eg, limited mouth opening, malocclusion, facial deformity). The MISS incorporated diplopia and age in the items scored. However, the calculation and assignment of the items are not undertaken in the same way. These 2 scoring systems inherited the disadvantages of the AIS. Nevertheless, the AIS includes classifications of maxillofacial bone and soft tissue and, with continual revision of the AIS, the classification of maxillofacial injury has improved greatly.

Mandibular fractures usually affect occlusion and mouth opening, which will influence patients’ QoL. Some investigators have devised scoring systems
specifically for mandibular fractures, such as the severity of mandibular fractures \(^{11}\) and mandibular injuries. \(^{12}\) The items in these scoring systems were primarily based on the experience of experts in maxillofacial trauma and have usually been used in prospective studies. They might show advantages in scoring for mandibular fractures; however, they have not been validated in published studies.

A validated scoring system for all types of maxillofacial injury has been lacking. Thus, the experience and judgment of experts in maxillofacial trauma have become the reference standard used to evaluate maxillofacial injuries. However, which scoring system is more effective and in accordance with expert opinion has not been confirmed. Expert scoring is a method for quantifying the qualitative description of maxillofacial trauma. In the present study, we initially chose the items that needed to be quantified. We then selected 35 experts and requested them to grade the injuries of 28 patients. We then selected 4 scoring systems (NISS, FISS, MFISS, MISS) with reference to 3 aspects: the extent of comprehensiveness, extent of recognition by Chinese and overseas experts in maxillofacial trauma, and practicality. The scoring system that will be most consistent with the expert score has not been reported.

The Pearson correlation coefficients among the expert score, 4 scoring systems, cost of operation, and operation time showed the expert score had a good correlation with the cost of the operation and operation time (\(R = 0.769\) and \(R = 0.759\), respectively), reflecting the injury severity adequately. Hence, we could conclude that the expert score was reliable.

We found that all 4 scoring methods correlated positively with the expert score. The correlation coefficient between the MFISS and the expert score was the greatest (0.801). Bland-Altman plots \(^{13}\) were used to calculate the consistency of the 2 results. This analytical method adopts the “difference” and “ratio” to reflect the consistency of 2 results. The closer to 0 for the difference or to 1 for the ratio, the better the consistency of the 2 results. Compared with the linear correlation coefficient, the Bland-Altman analysis is more stable.

We standardized the score of the 4 scoring systems owing to the large differences in their score ranges. Next, we used the standardized scores to test their agreement with the expert score. The difference between the MFISS and the expert score was \(-0.537 \pm 1.96\), and the ratio was \(0.46 \pm 0.19\), showing that the agreement between the MFISS and expert score was better than that of the other 3 scoring systems.

![Figure 8](image8.png)  
**FIGURE 8.** Bland-Altman plot showing the ratio of the standardized Facial Injury Severity Scale (SFISS) and expert score.  

![Figure 9](image9.png)  
**FIGURE 9.** Bland-Altman plot showing the ratio of the standardized Maxillofacial Injury Severity Score (SMFISS) and expert score.  

![Figure 10](image10.png)  
**FIGURE 10.** Bland-Altman plot showing the ratio of the standardized Maxillofacial Injury Severity Score (SMISS) and expert score.  
In addition to injury to anatomic structures, functional and esthetic factors also play important parts in the assessment of injury severity. The MFISS includes 3 such items: limited mouth opening, malocclusion, and facial deformity. The MFISS appeared to be closer to the judgment of the experts than to the other scoring systems. Compared with the MFISS, the MISS squared the greatest AIS score, and diplopia and age were added. Diplopia usually occurs with severe injury and, in general, makes a small contribution to injury severity. Injury to the facial soft tissue is included in the FISS, but its classification has not been sufficiently detailed. Only if the length of the laceration is longer than 10 cm is 1 point assigned to the score. A laceration longer than 10 cm is not common in maxillofacial injuries, and the damage to function and appearance will be seriously underestimated if the laceration is longer than 10 cm, but the severity score is assigned just 1 point. In addition, the classification of bone fractures is not meticulous, and more improvement is needed. In the present study, the NISS was based on the AIS-2005 version. The classification of maxillofacial injury has been greatly improved and refined in recent years. Although functional and esthetic factors were not included, the relevance and consistency between the NISS and expert score was high in the present study.

Although the MFISS was more consistent with the expert score in the present study, the expert score has not been validated. The conclusion derived from it needs to be confirmed in a larger patient cohort.

In conclusion, compared with the 3 scoring systems, the correlation and agreement between the MFISS and expert score was the greatest. This finding suggests that the MFISS is more suitable for the scoring of maxillofacial injury. However, its effectiveness must be validated in a larger patient cohort.

Acknowledgment

We thank the 35 experts for their help in our study.

References