Title: Inter-Rater Reliability of Acne Scar Grading and Correlation with Three Dimensional Facial Modeling Based Scar Quantification

N. Foolad¹, N. Prakash¹, T. Petukhova¹, V.Y. Shi¹, V. R. Sharon¹, L. O’Brecht², I.A. Ali¹, S. Feldstein³, J. Halls⁴, Q. Wang⁵ C.S. Li⁶, R. K. Sivamani¹

1. Department of Dermatology, University of California – Davis, Sacramento, CA USA
2. School of Medicine, George Washington University, Washington, DC USA
3. School of Medicine, University of California – San Diego, San Diego, CA USA
4. School of Medicine, Louisiana State University, Baton Rouge, LA USA
5. Department of Statistics, University of California - Davis, Davis, CA USA
6. Division of Biostatistics, Department of Public Health, University of California – Davis, Sacramento, CA USA

Correspondence: Raja K. Sivamani, Department of Dermatology, 3301 C Street, Suite 1400, Sacramento, CA 95816 USA, Phone: 916-703-5145, Fax: 916-442-5702, E-mail: rksivamani@ucdavis.edu.

Word count: 800, Table: 1 Figure: 1

Conflict of interest: None declared

Keywords: acne, scarring, grading, computer, imaging
DEAR EDITOR, Acne is the most common skin condition in the United States.(1) An estimated 40-50 million Americans have acne. Acne occurs more frequently among teenagers and young adults, although adults can get acne as well. Scarring is a common occurrence among individuals with acne. It has been estimated that facial scarring occurs in 95% of individuals with acne.(2) A key factor in planning treatment modalities for scar revision is objectively assessing scars by number and morphology.(3, 4) Goodman et al. constructed and reported a quantitative postacne scarring global severity scale.(5) With a score range of 0-84, this system resulted in a more objective evaluation that depended on scar type, count, and severity. There were a total of four graders in the study by Goodman et al. composed of physicians and nurses and they reported that their scoring system was reproducible regardless of medical background.(5) However, no inter-rater agreement statistics was reported.

In our study, we evaluated the Goodman et al. scar grading system with photographs taken by a 3D facial modeling and measurement photography device show in Figure 1 (Clarity™ 3D Research Ti System, BrighTex Bio-Photonics, San Jose, CA). The main objective was to assess the reproducibility of the postacne scarring global severity scale. Secondly, we measured the accuracy of this grading scale by comparing our graders’ median scores to a computer generated total scar volume calculation based on the depth and the area of involvement of the scars. The UC Davis Institutional Review Board approved this research protocol. All subjects provided written informed consent prior to participation.

Before the grading session, nine graders (consisting of medical students, dermatology residents, and board-certified dermatologists) underwent a group training session on the Goodman postacne scarring global severity scale. This included a training set that was administered to all
graders together to ensure similar training for all. Thereafter, the graders independently graded ten facial images of acne scarring. The interclass correlation coefficients were calculated for the medical students, dermatology residents, and the board-certified dermatologists. The imaging based total scar volume was correlated with the scar grading by the board-certified dermatologists by the Spearman’s rank-order correlation coefficient. All analyses were performed with SAS v9.4 (SAS Institute Inc., Cary, NC, USA).

The inter-rater agreeability differed by level of dermatology training. For medical students the ICC was 0.46 (95% CI: 0.14, 0.70), for dermatology residents the ICC was 0.62 (95% CI: 0.27, 0.82), and for board-certified dermatologists the ICC was 0.89 (95% CI: 0.66, 0.97). This suggests that the reliability of the Goodman scar grading system is dependent on dermatology training level.

The clinical grading scale was then correlated to a quantitative scar volume measure through 3D facial modeling and measurement technology. The computer analysis generated the total scar volume for each subject based on surface area and depth of surface defects on the face including the forehead, right and left cheeks, nose, and chin. When comparing the volume scores with the clinical grading by the board-certified dermatologists, there was a statistically significant positive correlation (0.76, p=0.01) as shown in Figure 1C.

In conclusion, it appears that level of dermatology training plays a role in the reliability of the Goodman scar grading system. Additionally, with the increasing development and role of facial imaging and modeling technology in medicine, there may be a way to incorporate quantitative computer-based facial analysis into research studies and patient treatment plans.
Acknowledgement
The project described was supported by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH), through grant #UL1 TR000002.

Table 1. Interclass Correlation Coefficient for Acne Scar Grading

<table>
<thead>
<tr>
<th>Rater</th>
<th>Interclass Correlation Coefficient and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Board Certified Dermatologists</td>
<td>0.89 (0.66, 0.97)</td>
</tr>
<tr>
<td>Dermatology Resident</td>
<td>0.62 (0.27, 0.82)</td>
</tr>
<tr>
<td>Medical Student</td>
<td>0.46 (0.136, 0.695)</td>
</tr>
</tbody>
</table>

Figure Legend

Figure 1. A high resolution facial photograph of a patient with acne scarring of the right cheek (A) and left cheek (B). (C) Correlation of board-certified dermatologist postacne scarring average score with computer generated total scar volume. The Spearman’s rank order correlation coefficient was 0.76 (p = 0.01).

References