Predicting Treatment Success after Scarf Osteotomy for Hallux Valgus using The American Orthopedic Foot and Ankle Society and Short Form Health Survey Scores

Meng Zhu, Jerry Y Chen, Terry HL Teo, Kevin OT Koo, Inderjeet S Rikhraj

ABSTRACT

Aim: Various hallux valgus corrective surgeries, including Scarf osteotomy, have demonstrated improvement in patients’ forefoot function and quality of life. However, no threshold values of these measures have been reported to define the success of surgery from the patients’ perspective. This study aims to define treatment success threshold values of The American Orthopedic Foot and Ankle Society (AOFAS)-Health Management Information System (HMIS) and Short Form (36) Health Survey (SF-36) Scores for patients underwent Scarf osteotomy (SO) for hallux valgus.

Materials and methods: A retrospective analysis of patients who underwent scarf osteotomy between 2007 and 2013 was conducted. Patients were evaluated for AOFAS-HMIS, SF-36 score, satisfaction, and pain score at various time points. Treatment success was defined as a significant improvement in pain and satisfaction with surgery. Using receiver operating characteristic (ROC) analysis, threshold score for treatment success was defined as the cut-off value providing the largest sum of sensitivity and specificity.

Results: A total of 360 and 345 patients completed assessments at 6-month and 2-year after surgery, with a success rate of 70.0% and 77.4%, respectively. The AOFAS-HMIS and physical component score (PCS) improved significantly at both postoperative time points as compared to preoperatively. The ROC analyses revealed the excellent predictive value of AOFAS-HMIS for treatment success (AUC = 0.840 and 0.835).

Conclusion: The American Orthopedic Foot and Ankle Society-Health Management Information System of 84 and above at 6-month, or 89 and above at 2-year after surgery can be used to define treatment success of SO for hallux valgus. An increasing higher threshold is required to define treatment success with time.

Keywords: Hallux valgus, Patient-reported outcome, Scarf osteotomy, The American Orthopedic Foot and Ankle Society-Health Management Information System, Threshold score, Treatment success.

How to cite this article: Zhu M, Chen JY, Teo THL, Koo KOT, Rikhraj IS. Predicting Treatment Success after Scarf Osteotomy for Hallux Valgus using The American Orthopedic Foot and Ankle Society and Short Form Health Survey Scores. J Foot Ankle Surg (Asia Pacific) 2018;5(2):59-63.

Source of support: Nil

Conflict of interest: None

INTRODUCTION

Hallux valgus is a common chronic forefoot condition that affects the function of foot and quality of life (QOL). Studies have demonstrated how various types of surgeries, including Scarf osteotomy, can improve the function and QOL through a combination of patient-reported outcome measures (PROMs) and different clinical parameters. While surgeons and healthcare professionals can monitor the changes in these scores for the progression of the condition and treatment response, there exists no threshold value to define treatment success in the patients’ perspective.

The AOFAS score is the most frequently used measurement tool for assessing the function of foot and ankle in research, and the AOFAS-HMIS has been shown to be reliable, valid, and responsive in clinical use. The Short Form 36-Item Health Survey (SF-36, RAND Health, Santa Monica, California) is a widely-employed health-related QoL evaluation method. Both AOFAS-HMIS and SF-36 score are the primary outcome measures used in the authors’ institution, similar in many others, for observing treatment response. Therefore, the objective of this study is to define the threshold scores of AOFAS-HMIS and SF-36 for successful treatment of hallux valgus using Scarf osteotomy.

MATERIALS AND METHODS

This study was approved by the hospital’s ethics committee and carried out in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. Written informed consent was obtained from all the patients.

Patients who underwent scarf osteotomy for correcting hallux valgus deformity at the foot and ankle division of a tertiary hospital during 2007 and 2013 were included. Relief of pain and correction of deformity were the indications for surgery. The patients were followed-up for 2 years.
years. They were assessed by independent healthcare professionals preoperatively, at 6-month and 2-year postoperatively, for forefoot function, QoL, and patient satisfaction. The AOFAS-HMIS was used to evaluate the forefoot function, which combines both PROMs and objective clinical evaluations. Thisscale assigned 40, 45, and 15 points to pain, function, and alignment respectively to give a maximally possible score of 100 points. The function component assessed includes daily and recreational activity limitation, footwear requirement, the presence of callus, as well as metatarsophalangeal and interphalangeal joint motion and stability.

Patient QOL was assessed by SF-36, which consists of eight subscales: physical functioning, physical role, bodily pain, general health, vitality, social functioning, emotional role, and mental health. The individual scores were compounded into two higher-order summary scores, the PCS and mental component score (MCS), based on the most highly correlated subscales to simplify analyses without substantial loss of information.16 These two summary scores have good validity in discriminating among different clinical groups, along with high test-retest reliability when used in a general population.17,18

During the follow-up assessment, the patients were assessed whether they are satisfied with the surgery. Besides, the pain was also rated on a 0 to 10 numerical scale, with 0 being no pain and 10 being the worst possible pain.19 Treatment success was defined as satisfaction with surgery, as well as pain-free or significant improvement of pain at the follow-up. Significant improvement in pain was defined as a reduction of pain score of either no less than two points or at least 30% of baseline score, whichever is higher.20

Statistical Analysis

Statistical analyses were carried out using ROC analyses in Statistical Package for the Social Sciences (SPSS) 21.0 (IBM, Armonk, NY, USA). The outcome measures, namely AOFAS-HMIS, PCS, MCS, and their respective changes at the two postoperative time points, were test-das predictors, while the defined treatment success was set as the criterion. An area under curve (AUC) between 0.80 and 0.90 indicates acceptable discrimination and an AUC above 0.90 indicates excellent discrimination.21

The most significant findings of the study are that AOFAS-HMIS is an excellent predictor of treatment success was defined as the cut-off value that provides the largest sum of sensitivity and specificity.22 Analyses were performed separately for both 6-month and 2-year follow-up for comparison.

RESULTS

Out of 364 cases included, a total of 360 (99%) and 345 (95%) cases were included and evaluated at 6-month and 2-year follow-up, correspondingly. Table 1 shows the demographics, preoperative PROMs as well as PROMs at the two follow-up time points. Significant improvements were noted in the AOFAS-HMIS (81 vs. 58, p < 0.001; and 87 vs. 58, p < 0.001) and PCS (49 vs. 46, p < 0.001; and 50 vs. 46, p < 0.001) at 6-month and 2-year follow-up.

Overall, 78.1% of patients had successful hallux valgus surgery at 6-month and 79.1% of patients at 2-year. Details on the fulfillment of the treatment success criteria are presented in Table 2. The ROC analyses demonstrated excellent prediction for treatment success using AOFAS-HMIS at both follow-up time points (AUC = 0.840 and 0.835), as shown in Figure 1. Acceptable AUC values were also observed if changes in the scale were used as predictors (AUC = 0.716 and 0.708). However, PCS, MCS, or their respective changes were not sensitive enough to predict treatment success (ROC curves for using PCS and MCS as predictors were not shown). Table 2 also summarizes the AUC values with 95% confidence range for each outcome measure used as predictors.

An AOFAS-HMIS of 84 and above at 6-month predicts successful treatment of hallux valgus with a sensitivity of 77% and a specificity of 77%, and a score of 89 and above at 2-year has a sensitivity of 73% and a specificity of 82%.

DISCUSSION

The threshold score of each outcome measure for predicting treatment success was defined as the cut-off value that provides the largest sum of sensitivity and specificity.22

### Table 1: Patient demographics and clinical outcome scores

<table>
<thead>
<tr>
<th></th>
<th>Pre-op (N = 364)</th>
<th>6-Month (N = 360)</th>
<th>2-Year (N = 345)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, male/female</td>
<td>25/339</td>
<td>25/335</td>
<td>23/322</td>
</tr>
<tr>
<td>Side, left/right</td>
<td>191/173</td>
<td>190/170</td>
<td>182/163</td>
</tr>
<tr>
<td>AOFAS-HMIS</td>
<td>58 ± 17</td>
<td>81 ± 16</td>
<td>87 ± 13</td>
</tr>
<tr>
<td>PCS</td>
<td>46 ± 4</td>
<td>49 ± 10</td>
<td>50 ± 8</td>
</tr>
<tr>
<td>MCS</td>
<td>55 ± 10</td>
<td>55 ± 10</td>
<td>55 ± 11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>6-month (N = 360)</th>
<th>2-year (N = 345)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment success</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction criteria</td>
<td>281 (78.1%)</td>
<td>273 (79.1%)</td>
</tr>
<tr>
<td>Pain criteria</td>
<td>304 (84.4%)</td>
<td>318 (92.2%)</td>
</tr>
<tr>
<td>Treatment success</td>
<td>252 (70.0%)</td>
<td>267 (77.4%)</td>
</tr>
<tr>
<td>AUC</td>
<td>0.840 (0.795–0.885)</td>
<td>0.835 (0.782–0.888)</td>
</tr>
<tr>
<td>Change in AOFAS-HMIS</td>
<td>0.716 (0.654–0.738)</td>
<td>0.708 (0.638–0.779)</td>
</tr>
<tr>
<td>PCS</td>
<td>0.587 (0.518–0.657)</td>
<td>0.643 (0.573–0.714)</td>
</tr>
<tr>
<td>Change in PCS</td>
<td>0.605 (0.537–0.674)</td>
<td>0.574 (0.501–0.646)</td>
</tr>
<tr>
<td>MCS</td>
<td>0.679 (0.620–0.738)</td>
<td>0.617 (0.544–0.690)</td>
</tr>
<tr>
<td>Change in MCS</td>
<td>0.578 (0.512–0.643)</td>
<td>0.627 (0.553–0.702)</td>
</tr>
</tbody>
</table>
success in hallux valgus correction using scarf osteotomy, and it can be used to interpret treatment outcomes both individually and in a cohort.

Hallux valgus is a common and complex deformity involving the first ray. Various clinical scores have been developed and validated to assess QOL and forefoot-specific function to guide management.\textsuperscript{12,13,23-26} They give the convenience for surgeons to observe the score changes in determining treatment response. However, treatment success cannot be accurately derived from the quantitative changes alone. Firstly, the change may be statistically meaningful, but clinically meaningless. Thus, minimal clinically important difference (MCID) has been proposed to be the smallest change in outcome scores which patients perceive as beneficial and meaningful.\textsuperscript{27, 28} While achieving the MCID is crucial for defining clinically meaningful improvement, it may not be considered as treatment success if the patient expectation is high. Secondly, the baseline score may affect the significance of the same magnitude of score change, with high baseline score requiring a possibly smaller magnitude of change to be construed as treatment success. Therefore, both absolute scores and their changes are considered as potential predictors in this study.

There is sparsity in the literature regarding the definition of treatment success in the context of hallux valgus corrective surgeries. Function scores and radiographic measurements are most commonly used in analyses and comparisons of surgical outcomes of hallux valgus correct surgeries. No definitive threshold values are provided in either absolute score or change in score to define treatment success. In addition, Schneider and Knahr found that using foot-specific clinical scores only partially addressed the important factors influencing the outcome of surgery, such as correction of footwear problems, alleviation of pain, and restoration of adequate walking.\textsuperscript{30} Criteria including pain and patient satisfaction are used in this study to define treatment success. This approach is more patient-centered and avoids physiological variables measured clinically or radiographically which may not be crucial in the patients’ perspective. Only patients underwent scarf osteotomy procedures are included in the study to minimize heterogeneity.

Hunt and Hurwit reported that a total of 139 unique clinical outcome scale being used in the literature on foot and ankle topics during the period of 2002 to 2011, and the AOFAS scales were the most commonly used PROMs for evaluating foot and ankle conditions.\textsuperscript{9} As AOFAS scales continue to gain popularity, the definitions of treatment success threshold scores for various foot and ankle procedures become more imminent. No similar definitions of treatment success threshold scores have been reported for PROMs in hallux valgus corrective surgeries. This study, the first of its kind, defines the threshold value of AOFAS-HMIS for treatment success of hallux valgus correction using scarf osteotomy.

Patient QOL is now recognized as one of the most important outcomes of surgery. Although there is a significant improvement in PCS postoperatively, the postoperative physical health status and improvement in physical health status were not meaningfully responsive for predicting treatment success. Pain relief is a significant contributor to the improvement of physical health QOL in hallux valgus,\textsuperscript{30} correlating to treatment success. Other components in PCS, such as role physical and general health, may contribute insignificantly to treatment success, diluting the usefulness of the summary scores. The limitation of SF-36 in predicting postoperative improvement was also elucidated by others.\textsuperscript{31}

There are a few limitations in this study. Firstly, there is a lack of radiographic assessment in determining treatment success. Studies have shown that the degree of

---

**Figures**

**Figures 1A and B**: Receiver operating characteristic (ROC) curve using AOFAS Hallux MTP-IP scale and its changes as predictors and treatment success as the criterion.
deformity and the amount of correction did not influence the outcome.\textsuperscript{1,2,33} Furthermore, patients may have considered the cosmetic outcomes in the reported satisfaction score subjectively. Secondly, different study population characteristics (age, gender, and follow-up duration) and different surgical procedures may give different threshold scores. The relatively large sample size could partially reduce the bias; however, the reported values should be interpreted as a guide rather than definition.

In conclusion, AOFAS-HMIS has excellent predictive value in defining treatment success in hallux valgus correction using SO, facilitating interpretation of treatment outcomes. The threshold score for treatment success varies with time, requiring a higher score to constitute treatment success at the later time point.

REFERENCES


27. King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert review of