Objective Structured Assessment of Technical Skills
in Elliptical Excision Repair of Senior Dermatology Residents
A Multirater, Blinded Study of Operating Room Video Recordings

Murad Alam, MD, MSCI; Michael Nodzenski, BA; Simon Yoo, MD; Emily Poon, PhD; Diana Bolotin, MD, PhD

Importance
Surgical training in dermatology residencies has increased, and there is growing interest in measuring resident competence more precisely. This study applies the use of the objective structured assessment of technical skills (OSATS) to measure competence in dermatologic surgery.

Objective
To evaluate the utility of OSATS as a tool for measuring surgery skills during dermatology residency training.

Design, Setting, and Participants
Multirater, blinded review and ratings of taped video recordings of elliptical excisions performed by senior dermatology residents applying for procedural or Mohs surgery fellowships.

Main Outcomes and Measures
Ratings on a specific OSATS measure, including global rating scale and task checklist.

Results
Twelve videos, representing approximately 20% of fellowship applicants during 2009-2010, were rated. Raters agreed on 272 of 288 subscore ratings (94.4%). Mean global ratings were 4 or 5 for all categories, except time and motion, which had a rating of 3. Task checklist ratings had a mean of 14.5 and a mode of 16 (perfect score), with eversion least often performed successfully. No association was found between a resident’s scores and the number of surgical rotation months or between scores and the number of Mohs surgeons at the home institution.

Conclusions and Relevance
Senior dermatology residents preparing for surgery fellowships are highly skilled in performing elliptical excisions and bilayered repairs. The OSATS appears useful and reliable for the evaluation of dermatologic surgery skills.

Published online March 12, 2014.

Dermatologic surgery has been an integral part of dermatology residency training for decades. Fortunately, dermatology residents receive more than adequate training in surgery, with continued growth in the quantity and breadth of procedures in which they participate.

Traditionally, residency training in basic suture and excision techniques has been conveyed via laboratory sessions with pig feet and gradually increasing responsibility in a traditional apprenticeship model. Faculty evaluations of resident competence in surgery are usually subjective, despite the attempts of some to develop and pilot specific criteria for basic skills.

In recent years, there has been an emerging consensus in other surgical specialties that evaluation of surgical skills in trainees can be performed systematically and relatively objectively. Potential benefits of such evaluations include the ability to precisely delineate competence and, thus, to more successfully intervene to improve deficient skills. Among the methods now used by surgical training programs to assess skill quantitatively are procedure-specific checklists, global ratings scales, objective structured assessment of technical skills (OSATS), motion analysis, virtual reality simulators, and video assessments.

The OSATS was one of the earliest methods developed for objective assessment of skills and has been most extensively evaluated by teachers of various surgical specialties. Composed of both a global rating scale and a procedure-specific checklist, OSATS addresses the concern that a unidimensional tool may be insufficient to accurately depict procedural competence. Originally validated 17 years ago (in 1997)
in comparison to expert faculty assessments of trainees’ competence, OSATS was used in early studies to assess trainees in bench or laboratory exercises on simulated patients. More recently, OSATS has been applied to evaluating competence in the operating room. The constituent measures of OSATS are modified for different types of surgical procedures and then validated before use.

The purpose of this investigation was to test the utility of OSATS for assessing the surgical competence of senior dermatology residents. The key skill that was tested was the ability to repair an elliptical excision with a bilayered closure.

Methods

Study Design

We performed multirater, blinded reviews and ratings of video recordings of dermatology residents performing elliptical excisions on patients. This study was approved by the Northwestern University Institutional Review Board as exempt and no informed consent was required.

Elliptical Excision Videos

Videos were selected from recordings submitted by applicants as an optional part of their application to the Northwestern University procedural dermatology fellowship program during the 2009-2010 academic year. Applicants were asked to submit a video of themselves performing a bilayered elliptical excision on the trunk or extremities at their home institution. All recordings that were deemed viewable and evaluable by a prescreener (S.Y.) were included. The criteria for inclusion were adequate technical quality (eg, focus, stability, frame, and picture quality) of the video, continuous and complete time course from skin preparation to end of the bilayered closure, and ability to see the surgeon’s hands, instruments, and the patient treatment site through the entire procedure. Video lengths varied from 10 to 28 minutes.

Videos were deidentified in that the only visible skin surfaces were the operating residents’ hands and partial forearms; the local areas that contained the excision site on the patient, ranging in size from approximately 10 to 30 cm wide, depending on draping; and, in some cases, the hands and forearms of assistants who were blotting or applying traction. Because the videos were of small excisions on the trunk and extremities, distortion of adjacent structures was not a major concern, and exposure of large areas was not necessary to ascertain the quality of the excision.

OSATS Global Rating Scale and Task Checklist

Although scales and task checklists are specifically developed for particular procedures, suturing is a skill required by all surgical specialties, and preexisting global scales and checklists were available. These scales and checklists were used without modification because they had previously been successfully tested and because a standardized instrument would potentially allow comparison across studies.

Rating Protocol

The 2 dermatologic surgeon raters (M.A. and D.B.) viewed each video in its entirety first. Then each rater viewed the same video again while scoring the checklist. The global assessment was scored at the end of the second viewing. Each rater completed all the ratings within 2 days to minimize performance appraisal bias associated with evolving standards. The 2 sets of ratings were reconciled using forced agreement.

Demographic Information on Surgeon Trainees

All videos were deidentified before evaluation, but information was retained pertaining to the characteristics of the associated dermatology residency programs, including their location, number of full-time dermatologic surgeons, and the mean number of surgery rotation months per resident. In addition, the number of publications per trainee was recorded.

Statistical Analysis

Descriptive statistics included the locations and characteristics of the residency institutions of surgeon trainees, the mean numbers of Mohs surgeons and months of surgery rotation per institution, and the mean number of publications per trainee. For global and checklist scores, means, ranges, and medians were computed. Means were used to describe the relationship of overall performance scores on the global rating measure to the following: (1) mean overall technical skill score, (2) mean number of surgical rotation months, (3) mean number of full-time Mohs surgeons at the residency institution, and (4) mean number of publications per trainee.

Results

Twelve of 13 submitted videos were evaluable. The associated third-year dermatology residents were training at 7 private and 5 public institutions (Table 1) in various regions of the United States and Canada. Eleven of the 12 programs had at least 1 full-time Mohs surgeon on staff (Table 1), with a range of 0 to 5. Surgeon trainees differed moderately with regard to months of surgical rotations in residency, with a median of 4 months. Most surgeon trainees had few publications but some trainees had many.

Rater evaluations of surgeon trainees were similar, with only 16 disagreements of 288 individual ratings (5.6%), and given the size of the surgeon sample, no further evaluation of interrater reliability was undertaken. The most common and mean global ratings were 4 or 5 for all categories, except for time and motion, which was rated 3 (Table 2). Task checklist ratings were similarly high (Table 3 and Figure), with the most common score being a perfect 16, with a mean of 14.5, and with evasion being the task least often performed successfully.

No association was found between a surgeon trainee’s scores and the number of surgical rotation months in residency or between scores and the number of Mohs surgeons at the home institution (Table 4).
Discussion

In this study, OSATS was used successfully to gauge the quality of bilayered repairs after elliptical excision performed by a subset of third-year dermatology residents. The OSATS entails rating surgical performance on 2 different metrics: the global rating scale and the task checklist. For this study, the global scale considered 8 criteria ranked from 1 (very poor) to 5 (clearly superior). The task checklist lists 16 surgical steps in sequential order, with each classified as performed correctly or not and 16 being a perfect score. Reassuringly, 9 of the 12 participants scored high. On the global rating scale, no applicant received a score below 3 out of 5 (competent), and 75% received scores of 4 or higher. On the task checklist, the mean score was 14.5 out of 16, with half the trainees receiving perfect scores.

There were some areas of relative skills deficiency. On the global rating scale, surgeon trainees performed the poorest on the time and motion criteria. On the task checklist, one-third were unable to achieve proper apposition of the wound and eversion, and at least 2 trainees each had difficulty with needle handling, including correct needle insertion, entry and exit points, and distance between sutures. Conversely, almost all trainees displayed superior knowledge of the procedure, were able to select appropriate instruments and suture, and handled tissue gently.

Significantly, the 2 raters exhibited a high degree of agreement, suggesting that the measures used are a robust and reproducible assessment tool. On the other hand, the task checklist did not appear to offer the same rating precision as the global rating scale. For instance, the 7 surgeon trainees who received a 4 on overall performance on the global scale had
same mean task checklist scores (mean score, 15) as the 2 trainees scoring 5 on the global scale (Table 4). This finding suggests that the global rating scale was better able to discriminate among small differences in performance.

Contrary to expectation, there appeared to be no link between the quantity (ie, number of months of surgical rotations during residency) or quality (ie, number of full-time Mohs surgeons at the program) of dermatologic surgery training and trainee’s competence in bilayered repair. The number of publications per trainee was negatively correlated with performance, with higher-performing surgeons having fewer publications. It is possible that applicants with more publications may have spent time in research that limited or interrupted their clinical training.

Perhaps surprisingly, most of the residents rated in this study performed very well. Potential reasons include the following: a basic skill, namely, elliptical excision, was tested, and most incoming senior dermatology residents have mastered this skill; residents were self-selected to be those who had special interest in surgery and were planning on pursuing a surgical fellowship; there was self-selection regarding which residents applying to this fellowship chose to submit the optional video; residents who submitted videos may have sent in their best work; and reviewers grading videos were given latitude in interpreting written standards.

It is unclear why there was not a correlation between skill in closing wounds and the number of months in surgical rotation. Possible reasons are that some may have performed relatively more hands-on surgery and less observation or assisting during surgery months, significant surgical training (eg, surgery during Veterans Affairs rotations) may have occurred outside formal surgical rotations, the timing of surgical months within residency may have affected the skill levels observed in videos, or a larger number of resident videos may be needed to detect a statistically significant difference.

There are limitations to our data. Given the number of surgeon trainees evaluated, it is difficult to make statistically valid comparisons between prior education and competence. For instance, the association between competence and publications was skewed by a single poorly performing applicant with many publications, and a much larger study would be needed to validate any such association. For our sample, it appears that individual factors were more important than where the residents trained and how much surgery experience they received in their residency. Another limitation is that all the surgeon trainees assessed were planning on completing surgical fellowships. As such, they may have had a different level of preparedness and skill than the overall population of incoming senior dermatology residents. Potentially, they may have been better surgeons, seeking to further develop this area of strength during fellowship, or inferior surgeons, seeking to correct a perceived deficit through fellowship, or a combination thereof. Regarding the measures used in this study, various parameters were not quantitatively defined. For instance, rating of economy of motion did not include satisfying a requirement that the surgeon’s hands stay within a distinct area in the surgical field, that not more than a certain number of separate hand motions should occur, or that total time required should be less than a benchmark. Finally, although the definitive outcome for skill in closing wounds is the long-term scar result, photographs of such could not be obtained in this study due to federal and institutional review board privacy considerations. No patient information was recorded, and apart from certain demographic factors pertaining to the surgeon, videos were deidentified even regarding surgeon name.

Strengths of this study include the rigor of the methods. The entire repair procedures were recorded on video and then observed several times during rating, 2 raters were used, preexisting OSATS tools that have been successfully refined in general and plastic surgery were used, and the surgeries observed were not laboratory practice on surrogates but operating room closures on patients. That raters evaluated recordings that did not show the faces of the surgeons was another major strength. Live OSATS ratings can create the risk of not only information loss if the rater is temporarily distracted but also potential bias if the rater knows the trainee and prejudices his or her competence.

In general use in dermatology, OSATS would be an inexpensive, easy-to-administer tool. Raters would not be burdened with remembering key evaluation criteria but precisely guided through the process through scales and checklists. The cost of administration would be low, and no additional surgical time would be added. In common practice, live evaluations, if slightly less accurate, would be more convenient than videos. The OSATS could be used in residency training, fellowship training, and hands-on continuing medical education courses. Specific OSATS tools could be developed to measure laser skills, injection skills, and other important procedural competences.

Future attempts to assess resident surgical skills using the measures described could be improved by quantification of the benchmark elements used to define quality, with this possibly via a consensus process among experts; greater uniformity of videography, including lighting, resolution, angle, and distance to the surgical field; calibration of raters based on standardized videos of poor, acceptable, or superior closures; and involvement of all residents at a training program for the entire duration of training, thus allowing for initial pretests and retests over time, assessment of long-term follow-up photo-

Table 4. Technical Skill Checklist Scores by Overall Performance Rating

<table>
<thead>
<tr>
<th>Overall Performance Score</th>
<th>No. of Residents</th>
<th>Mean Overall Technical Skill Score (of 16)</th>
<th>Mean Surgical Rotations, mo</th>
<th>Mean No. of Mohs Surgeries at Residency Institution</th>
<th>Mean No. of Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>15</td>
<td>6.0</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>15</td>
<td>3.8</td>
<td>2.3</td>
<td>1.3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>13</td>
<td>8.0</td>
<td>2.0</td>
<td>3.0</td>
</tr>
</tbody>
</table>
graphs, and tracking of skill level in concert with direct measures of surgical experience, such as Accreditation Council for Graduate Medical Education (ACGME) surgical logs. Appropriately adapted, OSATS may be useful for assessment of surgical milestones as a part of the ACGME milestones project.

Conclusions

Overall, this study found that OSATS is a reliable and useful measure that can be adapted to objectively assess competence in dermatologic surgery. In addition, dermatology residents appear to have achieved mastery of basic surgical skills toward the end of residency. Clinical applications of these findings may include use of OSATS to identify and remediate residents who are having difficulty with skin surgery. Pre- and post-OSATS testing may also help residency programs evaluate their success at improving resident skill levels over time. As appropriate, new OSATS tools may be developed to specifically teach and evaluate other skin cancer or aesthetic procedures performed by dermatologic surgeons. With the use of OSATS, dermatologists have a simple, easy-to-administer, and inexpensive procedure evaluation method that can be used to augment existing evaluation approaches.

REFERENCES