Creating a Cross-Institutional Grading Scale for Temporal Bone Dissection

Dinah Wan, BS, **Gregory J. Wiet, MD**, **D. Bradley Welling, MD, PhD**, **Thomas Kerwin, BS**, and **Don Stredney, MA**

Ohio State University College of Medicine (D.W.), Department of Otolaryngology (G.J.W., D.B.W., D.S.), Department of Computer Science and Engineering (T.K.), Department of Biomedical Informatics (G.J.W., D.S.), The Ohio State University; and the Ohio Supercomputer Center (T.K., D.S.), Columbus, Ohio, U.S.A.

Abstract

Objectives/Hypothesis—There is increasing interest in objective assessment of surgeon competence. In the field of otolaryngology, several surgical training programs, including The Ohio State University, the University of Toronto, and Stanford University, have pursued standardized criteria to rate their trainees' performance in the initial steps of temporal bone dissection (complete mastoidectomy with facial recess approach). Although these assessment metrics require the completion of similar basic components integral to successful temporal bone dissection, certain listed criteria are unique to each institution. Our aim was to establish a more standardized set of criteria that can be used across different institutions to objectively assess temporal bone dissection. We translated these new criteria into automated metrics in our temporal bone dissection simulator to achieve even more objective grading of temporal bone dissections.

Study Design—Cross-sectional study/survey.

Methods—The temporal bone assessment criteria developed by each of the three aforementioned institutions were compiled into an all-encompassing scale. This compilation was sent out as an online survey to members of the American Neurotology Society and American Otological Society with instructions to rate the importance of each criterion.

Results—Criteria that were ranked by >70% of respondents as either "very important" or "important" were used to create the new, cross-institutional scale for the objective assessment of temporal bone dissection.

Conclusions—The new assessment scale and its eventual incorporation into the temporal bone surgical simulator will enhance the objectivity of currently existing methods to evaluate surgical performance across different institutions.

Keywords

Objective assessment; surgeon competence; temporal bone dissection; cross-institutional; grading scale; surgical simulator

INTRODUCTION

Objective assessment of surgeon competence, historically one of the weakest aspects of surgical training,[1] is recently surfacing as an area of high interest.[2] Current methods to assess physician competency have been an emphasis of the Accreditation Counsel for Graduate Medical Education (ACGME) with the development and implementation of the six core competencies used to assess progress of physicians in training.[3] Despite the comprehensive nature of the competency assessments, technical skill remains buried within the patient care competency domain without any specific details as to proper assessment. Current attempts at technical skills assessment include: case logs, in-training and post-training assessment reports, oral examinations, and in some institutions, objective assessment of technical skill (OSATS).[4] These methods suffer from questionable reliability, validity, and practicality as applied to otologic surgery up to this point. One major undertaking of the ACGME is the creation of objective criteria to measure surgical performance.

Several training institutions, including The Ohio State University (OSU), Stanford University, and the University of Toronto, have developed criteria for grading surgical competency in temporal bone dissections. The Welling Scale (WS1), created at the OSU Medical Center, is a 35-item binary (0, 1) grading instrument currently used to evaluate temporal bone dissection skills of otolaryngology residents based on final product analysis of completed dissections on cadaveric temporal bones within the laboratory setting.[5] It emphasizes identification of temporal bone landmarks more so than any particular surgical approach. The University of Toronto has developed three types of scoring scales for temporal bone dissections, including the binary (yes/no) Task-Based Checklist (TBC), the Final Product Analysis, and the Global Rating Scale (GRS), a 1 to 5 rating scale that considers more generic aspects of performance, such as respect for tissue, flow of operation, and familiarity with technique.[6] Stanford's group has published 20 metrics used to assess temporal bone dissection skills on the mastoidectomy surgical simulator.[7] These metrics include drilling technique, suctioning technique, volume of bone removal, facial nerve exposure, and drill force and velocities. Although each scale requires the completion of common basic components integral to the successful completion of a mastoidectomy, some criteria diverge in regard to each institution's preferred approach to temporal bone dissection. The goal of this study was to create a more universal scale for assessment of temporal bone dissection that minimizes institution-specific surgical idiosyncrasies for a particular standardized mastoid procedure (complete mastoidectomy with facial recess approach to the middle ear). Our long-term objective is to be able to fully integrate these metrics into our simulator to provide completely objective assessment and direct comparison to expert performance. Additionally, the metrics will be used as the basis for an automated curriculum that will provide active feedback to the user, assuring proficiency in technical skill.

MATERIALS AND METHODS

A comprehensive temporal bone dissection grading scale was created that included all elements from the following institutional grading scales: 1) the WS1 from OSU; 2) the TBC, GRS, and Final Product Analysis from the University of Toronto; and 3) the 20 metrics to assess performance on the mastoidectomy surgical simulator from Stanford University. This comprehensive scale was formatted as an online survey on www.surveymonkey.com and sent out through email to 190 members of the American Neurotology Society (ANS) and 132 members of the American Otological Society (AOS). Instructions were given for the survey participants to rate each criterion as either "very important," "important," "moderately important," "of little importance," or "unimportant" to the successful

completion of a temporal bone dissection (survey questions were presented using a 5-point Likert item). Space was allocated at the end of each section for the respondent to leave comments. This online survey can be viewed at: http://www.surveymonkey.-com/s.aspx?sm=r85O4qEp_2b3ryqs2U_2fJ_2bIZA_3d_3d.

Survey results and comments recorded online over a period of 6 weeks were analyzed to identify the criteria that received the highest score in terms of their importance and relevance in the assessment of temporal bone dissection.

RESULTS

Eighty-eight responses were attained from 190 ANS members and 132 AOS members for a total response rate of 27%. Table I lists each criteria and the percentage of respondents who ranked each criteria as either "very important" or "important" for performance on a basic mastoidectomy. To create the new, cross-institutional scale for the objective assessment of temporal bone dissections (Table II), we included all criteria that were ranked by >70% of participants as either "very important" or "important." A lower-end cutoff of 70% was chosen because it was the minimum percentage that adequately excluded most (but not all) criteria that were repeatedly commented to be vague, unnecessary, or arbitrary to the successful completion of the complete mastoidectomy in temporal bone dissection.

Some repeated comments from survey participants are quoted below:

- "Straightness of edge" is a vague term.

- "Depth of cavity" is an arbitrary term and different depths are "appropriate" depending on procedure, pathology, and anatomy of the specific temporal bone.

- "Slow velocity" is almost never used near critical structures; rather, a "lighter touch" is used to achieve similar effects.

- "Cells remaining on tegmen" is dependent on individual pathology and may be okay as long as the tegmen is adequately identified or if there is no dural injury.

- Facial nerve may not always need to be exposed or visualized, depending on case.

- Semicircular canals (SCCs) are rarely, if ever, blue-line because it poses unnecessary risk.

- Jugular bulb may or may not be skeletonized or identified depending on case.

- Carotid artery is usually not exposed in a mastoidectomy.

- "Maintains appropriate distance between drill and suction" does not apply for those who use irrigation on the drill rather than the suction/irrigation; further, some exceptions exist where the drill and suction may be positioned >2 cm apart.

- Distinctions between basic mastoidectomy and complete mastoidectomy need to be made, as different landmarks are expected to be identified or exposed in each

DISCUSSION

Before any valid, reliable, and practical assessment of technical skill can be performed, a valid, reliable, and practical set of metrics must be developed. Dauphinee and Wood-Dauphinee recently outlined a process for developing evidenced-based medical education that involves defining the parameters to be measured, measuring those parameters, and benchmarking those parameters to assess educational outcomes.[8] The work recorded above is a first step in trying to identify those metrics that are important to successful temporal

bone dissection. By subjecting those metrics to expert review, a type of validation of those metrics is established. Additional work to validate these metrics will include developing benchmarking processes and incorporating them into an assessment tool followed by administration of that tool to determine validity. Current methodologies of technical skills assessment, including OSATS are labor intensive, expensive, and require dedicated laboratories. These metrics are comprehensive and provide assessment of operation-specific maneuvers, global ratings, and final product analysis, which are all necessary to establish a proper, objective assessment of technical skill.[9]

It is the comprehensive nature of the above metrics that requires the development of a new tool for it to be practical to administer. A future goal is to translate the elements of this newly compiled scale into an automated assessment tool. Each criterion on the grading scale can be translated into a quantitative metric programmed into our interactive simulator for temporal bone dissection. The individual metric scores are output data generated by the simulator based on various elements occurring during the drilling process, such as drill speed, proximity to critical structures, and percentage of voxels removed. By integrating quantitative metrics as such, the simulator can serve as a self-contained assessment tool, minimizing the need for an expert to view and evaluate the simulation playback at a lower level of training. This will provide not only summative feedback to the user in the form of an assessment of technical skill, but also provide the basis of an automated curriculum for teaching temporal bone dissection with formative feedback throughout the training process. This process of continuous feedback has been shown to be integral for efficient learning.[10]

Various groups working with surgical simulators have undertaken the development of integrated metrics. The Minimally Invasive Surgical Trainer-Virtual Reality simulator, one of the first laparoscopic simulators created, integrated metrics such as the number of motions made, time to completion of procedure, number of errors, economy of movement, and length of path during each simulation.[11] Sewell and colleagues[7] at Stanford University have also developed various integrated metrics for their mastoidectomy simulator by logging data such as users' hand movements and haptic forces into a console that calculates chosen metrics. The metrics they measured include drilling technique, suctioning technique, volume of bone removal, facial nerve exposure, drill forces, and drill velocities. The scores obtained from these automated metrics correlated strongly ($P < .01$) with global scores assigned by expert raters.[7]

Many comments recorded from the survey reflect the differences between criteria deemed important in the educational setting versus the clinical setting. For example, the comment that it is somewhat dangerous to blue-line the semicircular canals clinically is completely appropriate; but it is also dangerous to never have blue-lined them in the laboratory to know what they look like so that further penetration can be avoided in a clinical encounter. Likewise, identifying the jugular bulb and carotid artery may not be necessary in many complete mastoidectomy procedures clinically; however, it is a valuable exercise in the lab to identify these structures to become familiar with the anatomy of the vascular structures of the mastoid. These differences between standards expected in the clinical versus the educational settings are important to keep in mind as we approach the creation of a standardized temporal bone grading scale.

It is the long-term goal of this project to determine similar yet broader metrics to assess surgical performance on the temporal bone dissection simulator developed at our institution. We have initiated this process by assessing the feasibility of quantifying each criterion of the new cross-institutional grading scale into an automated metric. Table II lists the 25 criteria from our newly proposed assessment scale and the mode of assessment required to translate each criterion into a quantifiable metric.

The methods for quantifying the criteria from the new grading scale can be generalized to several categories. Contextual criteria can change depending on what is supposed to be done during a specific step. For example, burr type and size are not defined absolutely, but the requirements can change based on which step in the procedure is being done at the time. Conversely, some metrics can be defined absolutely. The proximity and violation categories can be simply checked by comparing the position of the drill to the structure. Violation occurs when part of the volume corresponding to a defined structure is removed, whereas proximity cases can occur when the tool is near a structure. Technique criteria are those that are evident during the procedure but cannot always be seen in final product analysis. This would include such criteria as "maintains visibility of tool while removing bone." The identification category is a difficult one because portions of the bone not directly associated with any specific structure are involved. This category includes thinning areas near structures to identify vital anatomy for the purpose of localization and protection. We propose that these types of criteria can be analyzed by comparing final products of student bones to composite datasets from several mastoidectomies performed on the same bone by expert surgeons. However, there is still research to be done in this area.

Preliminary trials have shown that the temporal bone dissector simulator developed at our institution has reached a standard of development to be used in otology residency training curriculums.[12–16] Other current efforts include validating the efficacy of this simulator through a multi-institutional study.[17] Assurance that simulators such as this can offer an objective and accurate assessment of surgical skills is crucial before their full implementation into otolaryngology training programs. As noted by Kneebone in regard to medical education, simulator use in surgical training, though promising, currently "races far ahead of evaluation."[18] By translating the cross-institutional grading scale into evaluative metrics built into the simulator, we will certainly take the evaluative aspect of surgical simulators a considerable stride forward. These metrics will ultimately need to be benchmarked and run through multiple trials to validate the efficacy of the simulator as an assessment tool for surgical preparedness.

As our practice environment continues to be scrutinized both internally and externally, development, validation, and implementation of metrics in an objective manner are paramount. Eventually, with the input from many within the specialty, we hope this work extends into initial certification and maintenance of proficiency and any other element that requires competency assessment. It is incumbent upon us to do the due diligence to provide this before outside forces do it for us.

CONCLUSION

The newly proposed temporal bone dissection grading scale enhances the objectivity of the currently existing grading scales by accounting for differences in surgical training across institutions. By eventually translating the elements of this new scale into metrics for automated assessment in our temporal bone simulator, we can ultimately achieve even more objective scoring of temporal bone dissections. The integration of the new scale into our surgical simulator will ultimately allow novice surgeons to receive automated and efficient feedback on their performance in the context of a more universal and standardized assessment scale.

Acknowledgments

BIBLIOGRAPHY

1. Darzi A, Mackay S. Assessment of surgical competence. Qual Health Care. 2001; 10(suppl 2):ii64–ii69. [PubMed: 11700382]

2. Shah J, Darzi A. Surgical skill assessment: an ongoing debate. BJU Int. 2001; 88:655–660. [PubMed: 11890231]

3. General Competencies [computer program]. Version 1.3. September 28, 1999. Chicago, IL: ACGME; 2000. Accreditation Council for Graduate Medical Education (ACGME). http://www.acgme.org

4. Sidu RS, Grober ED, Musselman LJ, Reznick RK. Assessing competency in surgery: where to begin? Surgery. 2004; 135:6–20. [PubMed: 14694296]

5. Butler NN, Wiet GJ. Reliability of the Welling scale (WS1) for rating temporal bone dissection performance. Laryngoscope. 2007; 117:1803–1808. [PubMed: 17721407]

6. Zirkle M, Taplin MA, Anthony R, Dubrowski A. Objective assessment of temporal bone drilling skills. Ann Otol Rhinol Laryngol. 2007; 116:793–798. [PubMed: 18074662]

7. Sewell C, Morris D, Blevins NH, et al. Validating metrics for a mastoidectomy simulator. Stud Health Technol Inform. 2007:421–426. [PubMed: 17377316]

8. Dauphinee WD, Wood-Dauphinee S. The need for evidence in medical education: the development of best evidence medical education as an opportunity to inform, guide, and sustain medical education research. Acad Med. 2004; 79:925–930. [PubMed: 15383347]

9. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via and innovative "bench station" examination. Am J Surg. 1997; 173:226–230. [PubMed: 9124632]

10. Grancharov TP, Reznick RK. Teaching procedural skills. BMJ. 2008; 336:1129–1131. [PubMed: 18483056]

11. Gallagher AG, Richie K, McClure N, McGuigan J. Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. World J Surg. 2001; 25:1478–1483. [PubMed: 11760752]

12. Bryan, J.; Stredney, D.; Sessanna, D.; Wiet, GJ. Proceedings of IEEE Visualization 2001. Washington, DC: IEEE Computer Society; 2001. Virtual temporal bone dissection: a case study; p. 497-500.

13. Stredney D, Wiet GJ, Bryan J, et al. Temporal bone dissection simulation—an update. Stud Health Technol Inform. 2002; 70:507–513. [PubMed: 15458142]

14. Wiet GJ, Bryan J, Dodson E, et al. Virtual temporal bone dissection simulation. Otolaryngol Head Neck Surg. 2001; 125:191.

15. Wiet, GJ.; Stredney, D.; Sessanna, D.; Bryan, J. Volume-based temporal bone dissection simulator. Paper presented at: AAO-HNSF/ARO Research Forum, Annual Meeting of the American Academy of Otolaryngology–Head and Neck Surgery Foundation; September 9–12, 2001; Denver, CO.

16. Wiet GJ, Stredney D, Sessanna D, Bryan J, Welling DB, Schmalbrock P. Virtual temporal bone dissection: an interactive surgical simulator. Otolaryngol Head Neck Surg. 2002; 127:79–83. [PubMed: 12161735]

17. Wiet GJ, Rastatter JC, Bapna S, Packer M, Stredney D, Welling DB. Training otologic surgical skills through simulation— moving towards validation: a pilot study and lessons learned. J Grad Med Educ. 2009; 1:61–66. [PubMed: 21975708]

18. Kneebone R. Simulation in surgical training: educational issues and practical implications. Med Educ. 2003; 37:267–277. [PubMed: 12603766]

TABLE I

Survey Results—Temporal Bone Dissection (Mastoidectomy With Facial Recess Approach) Grading Criteria With Ranked Importance.

Criteria to Be Included in the New Cross-Institutional Scale*	% Participants That Ranked Criterion as "Very Important" or "Important"	Criteria to Be Excluded From the New Cross-Institutional Scale†	% Participants That Ranked Criterion as "Very Important" or "Important"
Maintains visibility while removing bone	100	Drill force reduced within 4 mm of facial nerve	69.8
Selects appropriate burr type and size	98.9	Identifies the digastric ridge	69.4
Antrum entered	97.9	Horizontal SCC skeletonized	69.4
No violation of facial nerve sheath	97.7	Maintains appropriate distance between drill and suction	66.3
Sigmoid sinus is not entered	96.5	Does not use excessive drill velocity near critical structures	64.7
Identifies tympanic segment of the facial nerve	96.5	Sinodural angle sharply defined	64.7
Does not drill on ossicle	93.1	No cells remain on EAC	63.8
Does not use excessive drill force near critical structures	93	Identifies the jugular bulb	62.8
Firm, low, good hand position and grip on drill	91.9	Drill force reduced within 4 mm of dura	62.3
Identifies the chorda tympani or stump	90.7	Identifies the facial nerve at the stylomastoid foramen	60.7
Drills with broad strokes	88.5	Digastric tendon followed to stylomastoid foramen	58.2
Drills in best direction (clear understanding of cutting edge)	88.4	Identifies the carotid artery in middle ear	58.1
Identifies the facial nerve at the cochle-ariform process	88.3	No cells remain on tegmen	57
Appropriate depth of cavity (cortex)	86	Drill force reduced within 4 mm of sigmoid	56.5
Canal wall up (EAC)	85.6	Drills with unidirectional strokes	55.9
No holes in the EAC	84.9	Identifies the endolymphatic sac transition to duct	53
Complete saucerization (cortex)	83	SCCs blue-lined without fenestra	45.8
Posterior canal wall thinned	82.4	Use of diamond burr within 2 mm of dura	45.4
Low frequency of drill 'jumps' ('jump' defined as drilling further than 1cm from previous spot)	81.6	Use of diamond burr within 2 mm of sigmoid	42
Identifies the facial nerve at the external genu	80.9	Posterior SCC skeletonized	41.6
Facial recess completely exposed (overlying bone sufficiently thinned so nerve can be seen, located, and safely avoided)	75.9	Total time on task <30 min	40.4
No holes in tegmen	75.6	Drill velocity slowed within 4 mm of facial nerve	37.2
Use of diamond burr within 2mm of facial nerve	74.5	Superior SCC skeletonized	36.5
No cells remain on sinodural angle	71.8	Use of diamond burr within 2 mm of SCCs	34.1
		Drill force reduced within 4 mm of SCCs	30.5
		Sigmoid sinus completely exposed	25.7
		Drill velocity slowed within 4 mm of dura	23.5
		Drill velocity slowed within 4 mm of sigmoid	23.2
		Skeletonization of jugular bulb	23

Criteria to Be Included in the New Cross-Institutional Scale[*]	% Participants That Ranked Criterion as "Very Important" or "Important"	Criteria to Be Excluded From the New Cross-Institutional Scale[†]	% Participants That Ranked Criterion as "Very Important" or "Important"
		Drill velocity slowed within 4 mm of SCCs	22.1
		Straightness of edges (cortex)	19.3
		Stapedial muscle dissected	10.3

[*] Ranked by >70% participants as "very important" or "important."

[†] Ranked by <70% participants as "very important" or "important."

SCC = semicircular canal; EAC = external auditory canal.

TABLE II

Quantification of New Grading Scale Criteria Into Evaluative Metrics for the Temporal Bone Dissection Simulator.

Criteria to Be Included in the New Cross-Institutional Scale	Importance Rating (%)	Mode of Assessment for Quantification of the Metric*
1. Maintains visibility (of tool) while removing bone	100	Technique, contextual
2. Selects appropriate burr type and size	98.9	Contextual
3. Antrum entered	97.9	Violation
4. No violation of facial nerve sheath	97.7	Violation
5. Sigmoid sinus is not entered	96.5	Violation
6. Identifies tympanic segment of the facial nerve	96.5	Identification
7. Does not drill on ossicle	93.1	Violation
8. Does not use excessive drill force near critical structures	93.0	Technique Proximity
9. Firm, low, good hand position and grip on drill	91.9	Technique
10. Identifies the chorda tympani or stump	90.7	Identification
11. Drills with broad strokes	88.5	Technique, contextual
12. Drills in best direction (clear understanding of cutting edge)	88.4	Technique, contextual
13. Identifies the facial nerve at the cochleariform process	88.3	Identification
14. Appropriate depth of cavity (cortex)	86.0	Identification
15. Canal wall up (EAC)	85.6	Violation
16. No holes in EAC	84.9	Violation
17. Complete saucerization (cortex)	83.0	Identification
18. Posterior canal wall thinned	82.4	Identification
19. Low frequency of drill ''jumps'' (jumps defined as drilling further than 1 cm from previous spot)	81.6	Technique
20. Identifies the facial nerve at the external genu	80.9	Identification
21. Facial recess completely exposed (overlying bone sufficiently thinned so nerve can be seen, located, and safely avoided)	75.9	Identification
22. No holes in tegmen	75.6	Violation
23. Use of diamond burr within 2 mm of facial nerve	74.5	Proximity, technique
24. No cells remain on sinodural angle	71.8	Identification

*
Definitions for metric quantification modes of assessment: Contextual = contingent on sequence (steps); Technique = speed, force, angle, etc.; Proximity = distance of tool to structure(s); Violation = % of structure (volume) removed; Identification = % correlated with (composite) expert removal.

EAC = external auditory canal.