Title: The Melbourne Mastoidectomy Scale: validation of an end-product dissection scale for cortical mastoidectomy

Running title: MMS development and validation

Authors

Benjamin James Talks [a †], Jesslyn Lamtara [a], Sudanthi Wijewickrema [a], Jean-Marc Gerard [a], Alistair Macleold Mitchell-Innes [a], Stephen O'Leary [a]

[a] Department of Surgery (Otolaryngology), University of Melbourne, Royal Victorian Eye and Ear Hospital, Level 5/ 32 Gisborne Street, East Melbourne, Victoria 3002, Australia.

† Medical School, University of Birmingham, Edgbaston, Birmingham, West Midlands, B15 2TT.

Corresponding Author: Mr Benjamin James Talks, Department of Surgery (Otolaryngology), University of Melbourne, Royal Victorian Eye and Ear Hospital, Level 5/ 32 Gisborne Street, East Melbourne, Victoria 3002, Australia.
Email: BXT437@student.bham.ac.uk

Acknowledgements: Benjamin Talks would like to thank the University of Birmingham for his Student Development Scholarship which supported his involvement in this project.

Conflict of Interest Statement: the authors have no potential conflicts of interest with respect to the research, authorship, and/ or publication of this article.

Data Sharing: the data that support the findings of this study are available from the corresponding author upon reasonable request.

Funding Statement: this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

MR BENJAMIN JAMES TALKS (Orcid ID : 0000-0003-4896-2324)

Article type : Original Manuscript

Abstract

Introduction

Cortical mastoidectomy is a common Otolaryngology procedure and represents a compulsory part of Otolaryngology training. As such, a specific validated assessment score is needed for the progression of competency-based training in this procedure. Although multiple temporal bone dissection scales have been developed, they have all been validated for advanced temporal bone dissection including posterior tympanotomy, rather than the task of cortical mastoidectomy.

Methods

The Melbourne Mastoidectomy Scale, a 20-item end-product dissection scale to assess cortical mastoidectomy, was developed. The scale was validated using dissections by 30 participants (10 novice, 10 intermediate and 10 expert) on a virtual reality temporal bone simulator. All dissections were assessed independently by three blinded graders. Additionally, all procedures were graded with an abbreviated Welling Scale by one grader.

Results

There was high inter-rater reliability between the three graders ($r = 0.9210$, $p < 0.0001$). There was a significant difference in scores between the three groups ($p < 0.0001$). Additionally, there was a large effect size between all three groups: the differences between the novice group and both the intermediate group ($p = 0.0119$, $\eta^2 = 0.2482$) and expert group ($p < 0.001$, $\eta^2 = 0.6356$) were significant. The difference between the intermediate group and expert group again had a large effect size ($\eta^2 = 0.3217$), but was not significant. The Melbourne Mastoidectomy Scale correlated well with an abbreviated Welling Scale ($r = 0.8485$, $p < 0.0001$).

Conclusion

The Melbourne Mastoidectomy Scale offers a validated score for use in the assessment of cortical mastoidectomy.

Key Words: cortical mastoidectomy; virtual reality; Otolaryngology; competency-based education; simulation training; educational measurement.

Key Points

- Cortical mastoidectomy is a foundational core competency in Otolaryngology.
- Competency based surgical training needs validated objective skills assessments.
- We developed a binary 20-item scale for assessment of cortical mastoidectomy.
- Validation was performed with 30 participants on a virtual reality simulator.
- The scale has high inter-rater reliability and can separate groups by skill level.

Introduction

Cortical mastoidectomy is a common operation for Otologists with a number of indications including chronic otitis media with or without cholesteatoma; it is also the initial step of cochlear implant surgery and various lateral skull base operations. In addition, general Otolaryngologists need to be able to carry out cortical mastoidectomy safely as an emergency procedure for acute mastoiditis. Subsequently, it is a compulsory procedure for Otolaryngology training in the United Kingdom (1). During a mastoidectomy, the surgeon comes into close proximity with a number of important structures: the middle fossa dura mater superiorly, the sigmoid sinus posteriorly, and the facial nerve, semicircular canals, and incus medially (2). Clearly, it is important for surgical trainees to gain competency in this procedure to avoid major complications and allow progression to more advanced temporal bone operations.

Until recently, surgical competency has been assessed through logbooks of surgical experience, written and oral exams, and informal observation of operating skill by supervisors as part of a surgical apprenticeship (3). Modern surgical training programmes are increasingly being built upon the principle of competency-based training (4), relying on validated scores to standardise the assessment of technical skills. At present, various scales have been developed for the assessment of temporal bone surgery (6–10), with the Welling Scale being the most widely used of these (6). All these current scores have been developed and validated for the assessment of canal-wall up cortical mastoidectomy with facial recess exposure (6–10), which represents an important but advanced milestone in Otolaryngology training (1). On the other hand, cortical mastoidectomy represents a key early surgical competency in Otology, and a skill that all Otolaryngologists irrespective of sub-specialty need to be able to master to manage acute mastoiditis, an uncommon but potentially life-threatening condition. At present a cortical mastoidectomy has no specific objective assessment scale to determine proficiency.

The majority of scales are designed to assess the final-product of temporal bone dissection (6–9), although some task-based checklists have also been developed to assess the dissection process (9,10). Despite providing more information about surgical technique, task-based checklists are considerably more time consuming than final product scores, requiring the assessor to watch the entire procedure.

Previous scales have been scored in a variety of ways ranging from Likert Scales (9), to more complex sums of positively weighted items for procedure completion and negatively weighted items for errors (8), to binary scoring systems (6,7). Validation of these scales has been predominantly performed using cadaveric temporal bone dissections by a small number of Otolaryngology registrars (6,8,9). Only one study has evaluated the validity of such assessment scales on virtual reality (VR) temporal bone simulators (11), using a modified version of the Welling Scale; the only assessment of scale validity in this study was interrater reliability for 34 novice registrars. The properties and validation work for previously published temporal bone final product assessment scales are summarised in table 1.

In this study, we developed a new end-product dissection scale, the Melbourne Mastoidectomy Scale (MMS), which is specifically tailored to the foundational Otolaryngology operation of cortical mastoidectomy, with the aim of providing an objective tool for assessing competency performing this core procedure. We then validated this scale as to its objectivity (inter-rater reliability), ability to differentiate between skill levels, and its correlation with a pre-existing means of cortical mastoidectomy assessment (the Welling Scale) using a VR simulator.

Methods

Ethical considerations

This study was approved by the Human Research Ethics Committee of the Royal Victorian Eye and Ear Hospital (HREC number: 19/1419HL). All participants provided signed consent.

Development of the Melbourne Mastoidectomy Scale

The MMS scale was developed as a consensus decision between two Otology consultants and an Otolaryngology researcher, with the goal of creating a simple, easy to use scale targeted specifically at cortical mastoidectomy and with the potential to be automated on VR simulators. The end-result was a 20-item end-product dissection score (Table 2). This scale was tailored to assess the foundational core competency of cortical mastoidectomy, which is required of all Otolaryngology trainees, with consideration of novice registrars learning the procedure. The scale was designed to be easy to use with a simple binary scoring system comprising only 20-items, each with clear definitions. Additionally, each item was defined in terms of volumes of bone removed or volumes of structures damaged to facilitate automation of the score on VR simulators.

Similar to the Welling Scale and CanadaWest Scale (6,7), the scale has a binary scoring system, which aims to minimise grader subjectivity to improve inter-rater reliability. The assessment criteria owe much to the original Welling Scale (6); however, only 20-items were included in the scale to reduce grader exhaustion associated with the 35-items of the original Welling Scale. Each item of the scale is clearly defined to help standardise its interpretation by different graders, the CanadaWest Scale is the only other scale to include similar definitions (7). Also in common with the CanadaWest Scale (7), point dependencies have been introduced to the scale to reduce the number of points users are awarded due to an incomplete dissection. For example, a point for not damaging the sigmoid sinus is only awarded if the sigmoid sinus has been identified during the dissection. Additionally, 4 items of the scale have been defined as major complications, which mark the dissection as unacceptable, irrespective of the total score achieved. This additional classification is necessary in addition to a numerical score to safely define a trainee as competent to perform a cortical mastoidectomy and to progress to more advanced temporal bone procedures.

The virtual reality simulator

Validation of the MMS was carried out on the University of Melbourne VR Temporal Bone Simulator (12). This simulator presents the user with three-dimensional (3D) virtual models of temporal bones, generated from microCT scans of human temporal bones. A haptic device, represented in the virtual operating space as a surgical drill, is used to interact with the temporal bones and provides tactile feedback such as resistance and vibrations (Figure 1). Although the simulator has several automated guidance features, such as proximity warnings when approaching an important structure, these were turned off for the purpose of this study.

Study participants

Study participants (n = 30) of three experience levels were recruited for the validation of the MMS: novice (n = 10), intermediate (n = 10), and expert (n = 10). Novice participants were University students with an interest in surgery, intermediate participants were Otolaryngology registrars training in Melbourne ranging from Australian Surgical Education Training program stage 1 (junior) to stage 5 (senior), and expert participants were consultant Otologists.

Study procedure

All participants were given time to familiarise themselves with the VR simulator before being asked to carry out a cortical mastoidectomy, defining MacEwan's triangle and identifying the middle fossa plate, sigmoid sinus, incus, and vertical section of the facial nerve. The novice participants were shown a 15-minute video tutorial on how to perform a cortical mastoidectomy before carrying out their dissection. All participants were presented with the same temporal bone on the VR simulator, which was intended to be an easy specimen with no abnormal or pathological anatomy.

Outcomes

A screen recording of the final product of each cortical mastoidectomy procedure was captured for later grading. Cortical mastoidectomies by all participants were graded with the MMS by three graders (two Consultant Otologists and one Otolaryngology researcher). Additionally, all procedures were graded with an abbreviated version of the Welling Scale, using only items relevant to the cortical mastoidectomy task used in this study (18/35 items; appendix A), by one grader (Otolaryngology researcher). All screen recordings were anonymised before assessment, blinding the graders to the identity and experience level of the participant.

Statistical methods

Statistical analysis was carried out using MATLAB R2019b (Mathworks, Natick, USA). Inter-rater reliability was assessed using the intra-class correlation coefficient (ICC) (13). A one-way ANOVA with post-hoc Tukey-Kramer tests for multiple comparisons between the three experience groups: novice, intermediate, and expert, was conducted to test for the ability to differentiate skill levels. Effect sizes between groups were calculated using partial eta-squared. Pearson's Correlation Coefficient was used to assess correlation of the MMS with an abbreviated Welling Scale. All statistical tests were performed at the level of alpha = 0.05. Effect sizes were interpreted as negligible, $\eta^2 < 0.01$; small, $0.01 \leq \eta^2 < 0.06$; moderate, $0.06 \leq \eta^2 < 0.14$; and large, $\eta^2 \geq 0.14$ (14,15).

Results

A total of 30 procedures were assessed independently by three graders. There was high interrater reliability between the three graders ($r = 0.9210$, $p < 0.0001$). Mean ± standard deviation MMS scores for the novice, intermediate, and expert groups were 10.7 ± 3.51, 14.4 ± 2.79, and 17.3 ± 1.28 respectively. There was a significant difference in MMS score between the three groups ($p < 0.0001$; figure 2). When looking at specific groups, novices were significantly worse than both intermediates ($p = 0.0119$) and experts ($p < 0.001$), but there was no significant difference between the intermediate and expert groups. However, large effect sizes were observed between all three groups (novices and intermediates: $\eta^2 = 0.2482$, novices and experts: $\eta^2 = 0.6356$, and intermediates and experts: $\eta^2 = 0.3217$).

On examination of individual points on the scale, there were large effect sizes between the three groups for 12 items, of which 10 were significant. When directly comparing the novice and expert groups, there was a large effect size again for 12 items, of which 9 were significant; when comparing the novice and intermediate group 7 of these items had a large effect size, of which 4 were significant. There was a large effect size for 5 items when comparing the intermediate and expert group, of which only 1 was significant (appendix B).

The MMS correlated well with the abbreviated Welling Scale used in this study ($r = 0.8485$, $p < 0.0001$, figure 3).

Discussion

Previous temporal bone assessment scales have been developed for advanced temporal bone dissections. This study developed and validated the first assessment scale explicitly for the core foundational competency of cortical mastoidectomy.

There was high inter-rater reliability for the MMS between the three graders, which was higher (6,11) or equivalent (7,8) to previously reported temporal bone assessment scales. Inter-rater reliability is a key feature of any assessment tool for surgical training, allowing it to be used in different centres by different assessors. Consistent usage of the scale by different graders is supported by the provision of a clear definition for each item. Only one other temporal bone assessment scale has similar descriptions for each item (7).

The ability of the MMS to distinguish between low and high quality dissections as performed by operators of different experience levels was demonstrated by the large effect size when comparing cortical mastoidectomy performance between each of the three groups. As expected, the novice group performed significantly worse than both the intermediate and expert groups. However, there was not a significant difference in performance between the intermediate and expert groups, which is not surprising given that cortical mastoidectomy is a relatively basic procedure in which Otolaryngology trainees would be expected to achieve competency in early in their training. Furthermore, since the intermediate group contained trainees across the full range of postgraduate training (stages 1-5 of the Australian Surgical Education Training program), we would expect a degree of overlap between the two groups. The large effect size of the difference between the intermediate and expert groups suggests that its lack of significance is likely to be attributable to an insufficient number of participants in the two groups to account for their overlap in ability.

When looking at specific items of the scale there was a large effect size for 12 items (10 significant) between the novice and the expert groups, for 7 items (4 significant) between the novice and intermediate groups, and for 5 items (1 significant) between the intermediate and expert groups. A substantial contribution to the differences seen between the novice group and the intermediate/ expert groups of surgeons were attributable to damage of important structures, representing 4 and 6 of the items with a large effect size between the two groups respectively. Whilst, 4 of the 5 items with a large effect size between the intermediate and expert groups represent inadequate removal of bone to expose landmarks rather than damage to important structures. Of note, novices scored significantly worse than both the intermediate and expert groups for four points with large effect sizes: opening of the antrum, identifying the incus, and avoiding damage of the ossicular chain and semicircular canals. These represent key stages of the cortical mastoidectomy procedure that novices need support completing, and fit with a previous study highlighting the significantly greater incidence of ossicular chain and labyrinth damage in novice versus experienced Otolaryngology trainees (7). Future studies to identify the stages of the cortical mastoidectomy procedure that cause the greatest difficulty to trainees at each stage of their training could be helpful to allow the development of a targeted task-specific VR surgical curriculum, involving extensive practice of the identified operative steps.

The Welling Scale is the most widely used of the current temporal bone scores. The MMS correlated well with the abbreviated Welling Scale used in this study (appendix A), demonstrating equivalence with the current standard of cortical mastoidectomy assessment. Of note, only 18 of the 35 points of the original Welling Scale were used in this study as the scale includes assessment of posterior tympanotomy and detailed anatomical dissection of the temporal bone, which was not the task being investigated in this study.

Clinical applicability of the study

The high objectivity of the scale, as evidenced by its high interrater reliability and its ability to separate high quality dissections from poorer quality ones performed by individuals with varying surgical experience makes it appropriate for use in cortical mastoidectomy competency assessments. The objective assessment of competency in cortical mastoidectomy would help determine when trainees are able to progress to more complex temporal bone procedures and allow effective revalidation of competency in the procedure by subspecialised Otolaryngologists, who all may be expected to perform a cortical mastoidectomy in the emergency context of life-threatening acute mastoiditis. The classification of four of the items of the scale as major complications is important to help determine individuals' safety performing a cortical mastoidectomy in addition to their numerical score.

Limitations of the study

A limitation of this study is that cortical mastoidectomy performance was only assessed on a VR simulator. It would be beneficial to additionally validate the scale for use on cadaveric temporal bones, other modes of simulation such as 3D printed plastic bones, and in the operating room. However, as an end-product dissection scale, we would expect the MMS to translate well to these other settings; although the environment changes for the operator, the end goal of the procedure stays the same and the only real change for the assessor is the varying appearance of anatomical structures inherent to each setting such as colour. As the MMS is specific to the task of cortical mastoidectomy, it is most relevant to junior surgical trainees, who would benefit most from practice on VR simulators, conserving the valuable resource of cadaveric bones for more senior trainees. Additionally, it is the only scale that has been robustly validated for assessing VR dissections. Although one previous study assessed VR dissections using a modified Welling Scale, the only validation they carried out for this score was inter-rater reliability for two graders, whilst the modification of the score itself was not validated (11). VR simulation presents an ideal platform for competency tests – able to both present a standardised dextrous task and record detailed metrics on performance of the task. Such work based competency assessments have been a compulsory part of surgical training in United Kingdom since 2007 and have shown validity in the assessment of Otolaryngology trainees (16). Another advantage of validating this score for use on VR simulators is the potential for its automation. The simple design of this scale using binary points and focussing on structure exposure or damage supports the feasibility of its automation using routinely collected simulator metrics such as volumes of bone drilled.

Conclusion

In conclusion, the MMS offers a validated temporal bone score for assessment of cortical mastoidectomy. As such, it could be of particular value to the early training of Otolaryngology registrars, who must gain competency in cortical mastoidectomy before progressing to more advanced temporal bone procedures, and subspecialised Otolaryngologists seeking to revalidate their competency in the procedure.

References

1. Intercollegiate Surgical Curriculum Programme. The Intercollegiate Surgical Curriculum. Educating the surgeons of the future: Otolaryngology [Internet]. 2017 [cited 2019 Jul 4]. Available from: https://www.iscp.ac.uk/curriculum/surgical/surgical_syllabus_list.aspx
2. Sanna M, Khrais T, Falcioni M, Russo A, Taibah A. The Temporal Bone: A manual for dissection and surgical approaches. Stuttgart; New York: Thieme; 2006.
3. Reznick R. Teaching and testing technical skills. Am J Surg. 1993;165:358–61.
4. Bhatti N, Cummings C. Viewpoint: Competence in surgical residency training: Defining and raising the bar. Acad Med. 2007;82:569–73.
5. Martin J, Regehr G, Reznick R, Macrae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997;84:273–8.
6. Butler N, Wiet G. Reliability of the Welling Scale (WS1) for rating temporal bone dissection performance. Laryngoscope. 2007;117:1803–8.
7. Pisa J, Gousseau M, Mowat S, Westerberg B, Unger B, Hochman J. Simplified summative temporal bone dissection scale demonstrates equivalence to existing measures. Ann Otol Rhinol Laryngol. 2018;127(1):51–8.
8. Mowry S, Woodson E, Gubbels S, Cargrae M, Hansen M. A simple assessment tool for evaluation for cadaveric temporal bone dissection. Laryngoscope. 2018;128:451–5.
9. Zirkle M, Taplin M, Anthony R, Dubrowski A. Objective assessment of temporal bone drilling skills. Ann Otol Rhinol Laryngol. 2007;116(11):793–8.
10. Laeeq K, Bhatti N, Carey J, Della Santina C, Limb C, Niparko J, et al. Pilot testing of an assessment tool for competency in mastoidectomy. Laryngoscope. 2009;119:2402–10.
11. Andersen S, Cayé-Thomasen P, Sørensen M. Mastoidectomy performance assessment

of virtual simulation training using final-product analysis. Laryngoscope. 2015;125:431–5.

12. O'Leary S, Hutchins M, Stevenson D, Gunn C, Krumpholz A, Kennedy G, et al. Validation of a networked virtual reality simulation of temporal bone surgery. Laryngoscope. 2008;118(6):1040–6.

13. McGraw K, Wong S. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1(1):30–46.

14. Richardson J. Eta squared and partial eta squared as measures of effect size in educational research. Educ Res Rev. 2011;6:135–47.

15. Cohen J. Chapter 8 - F tests on means in the analysis of variance and covariance. In: Statistical power analysis for the behavioural sciences. Academic Press; 1977. p. 273–406.

16. Awad Z, Hayden L, Robson A, Muthuswamy K, Tolley N. Reliability and validity of procedure-based assessments in otolaryngology training. Laryngoscope. 2015;125:1328–35.

Figure Legends

Figure 1. The University of Melbourne Virtual Reality Temporal Bone Simulator presents the user with a 3D image of a temporal bone model which can be interacted with using a virtual drill provided by the haptic arm.

Figure 2. There was a significant difference in Melbourne Mastoidectomy Scale score between the three groups ($p < 0.0001$).

Figure 3. The Melbourne Mastoidectomy Scale correlated well with the abbreviated Welling Scale used in this study ($r = 0.8485$, $p < 0.0001$).

Table 1. The structure and validation methods of the four current final product temporal bone dissection scales.

Scale name	Number of items	Scoring	Validation participants	Validation modality	Interrater reliability
University of Toronto	14	Five-step Likert Scale	19 registrars	Cadaveric bones (x19)	2 graders Moderate (r = 0.60)
University of Iowa	3 items for dissection completeness 20 items for errors	Worth 7 to 10 points (max +25) Worth -1 to -4 points (max -25)	30 registrars	Cadaveric bones (x30)	5 graders High (r = 0.883)
Welling Scale	35	Binary	12 registrars	Cadaveric bones (x21) and plastic bones (x5)	6 graders Moderate (κ = 0.584)
CanadaWest Scale	14 (4 dependent on prior item)	Binary	19 registrars	3D printed bones (x19)	4 graders High (κ= 0.858)

Table 2. The Melbourne Mastoidectomy Scale.

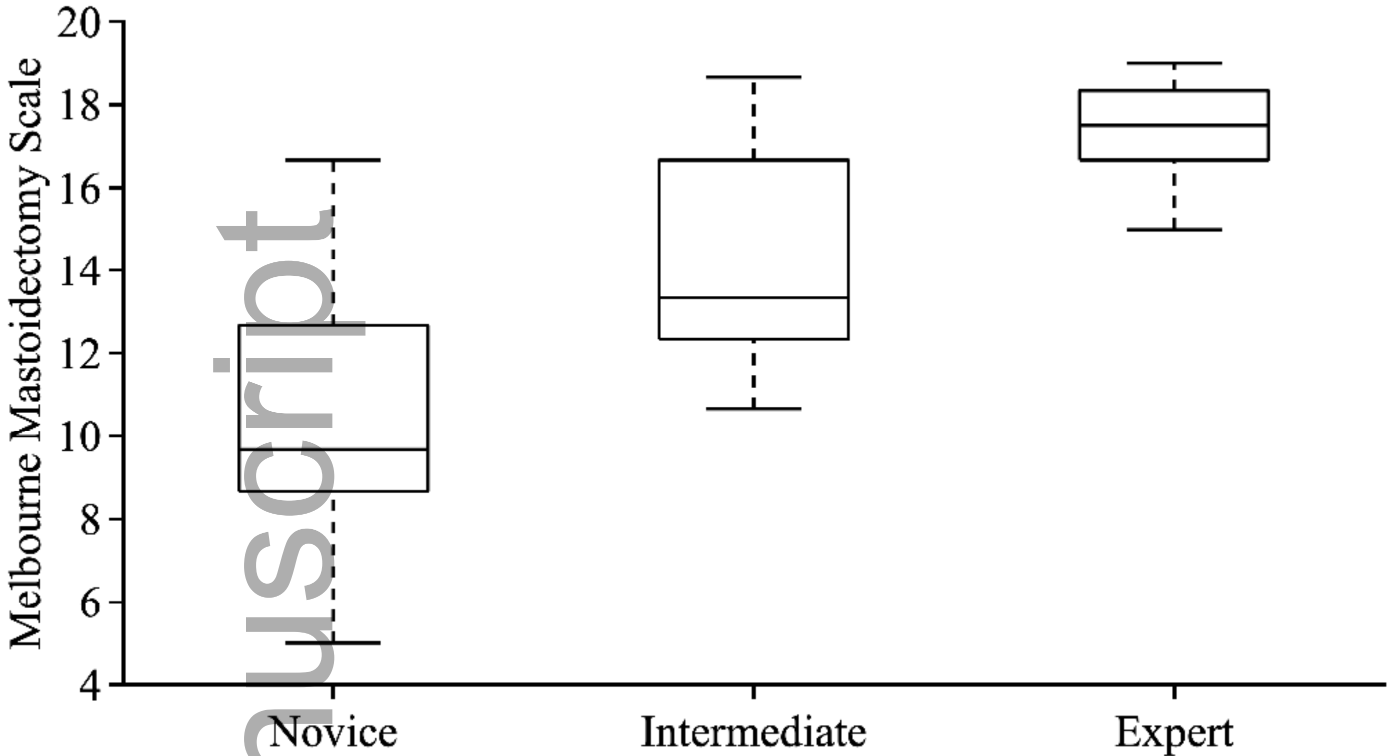
	Definition	Disagree	Agree
MacEwans Triangle defined as			
1. Temporal line	Cortex removed along the temporal line, delineating the superior limit of dissection.	0	1
2. Posterior external auditory canal wall	Cortex removed behind the posterior wall of the external auditory canal, defining the anterior limit of dissection.	0	1
3. Sigmoid sinus	Cortex removed over the suspected course of the sigmoid sinus, from the temporal line towards the mastoid tip, defining the posterior limit of dissection.	0	1
Middle fossa plate			
4. Identified	Partial exposure/clear identification of the middle fossa plate.	0	1
5. Adequately exposed [4]	Skeletonised middle fossa plate from sinodural angle to tegmen tympani without overhanging cortex.	0	1
6. Identified without minor damage [4]	No small holes in the middle fossa plate.	0	1
7. Identified without major damage [4] †	No large holes in the middle fossa plate or drilling of the underlying dura.	0	1
Sigmoid sinus			
8. Identified	Partial exposure/ clear identification of the sigmoid sinus.	0	1
9. Adequately exposed [8]	Skeletonised sigmoid sinus from sinodural angle towards mastoid tip, without overhanging cortex.	0	1
10. Identified without damage [8] †	No holes in the overlying bone or direct drilling of the sigmoid sinus.	0	1
11. Sinodural angle defined [8]	Sharp angle between the exposed sigmoid sinus and middle fossa plate.	0	1

External auditory canal			
12. Canal wall preserved	Grosely skeletonised external canal wall.	0	1
13. Posterior canal wall adequately thinned [12]	Precisely skeletonised external canal wall on at least 130 degrees.	0	1
14. Canal wall thinned with no holes [13]	No holes in the external canal wall.	0	1
Mastoid antrum			
15. Antrum opened	Drilling to open the mastoid antrum with exposure of lateral semi-circular canal.	0	1
16. Antrum opened with no damage of the semicircular canals [15] †	All the semicircular canals remain intact, with no holes.	0	1
17. Incus identified	The entire superior edge of short process of the incus is visible.	0	1
18. Incus identified without damage [17]	No drilling or disruption of the ossicular chain.	0	1
Facial nerve			
19. Vertical section identified	The vertical section of the facial nerve is visible.	0	1
20. Identified with no damage [19] †	No exposure of facial nerve sheath.	0	1
TOTAL SCORE		**/20**	

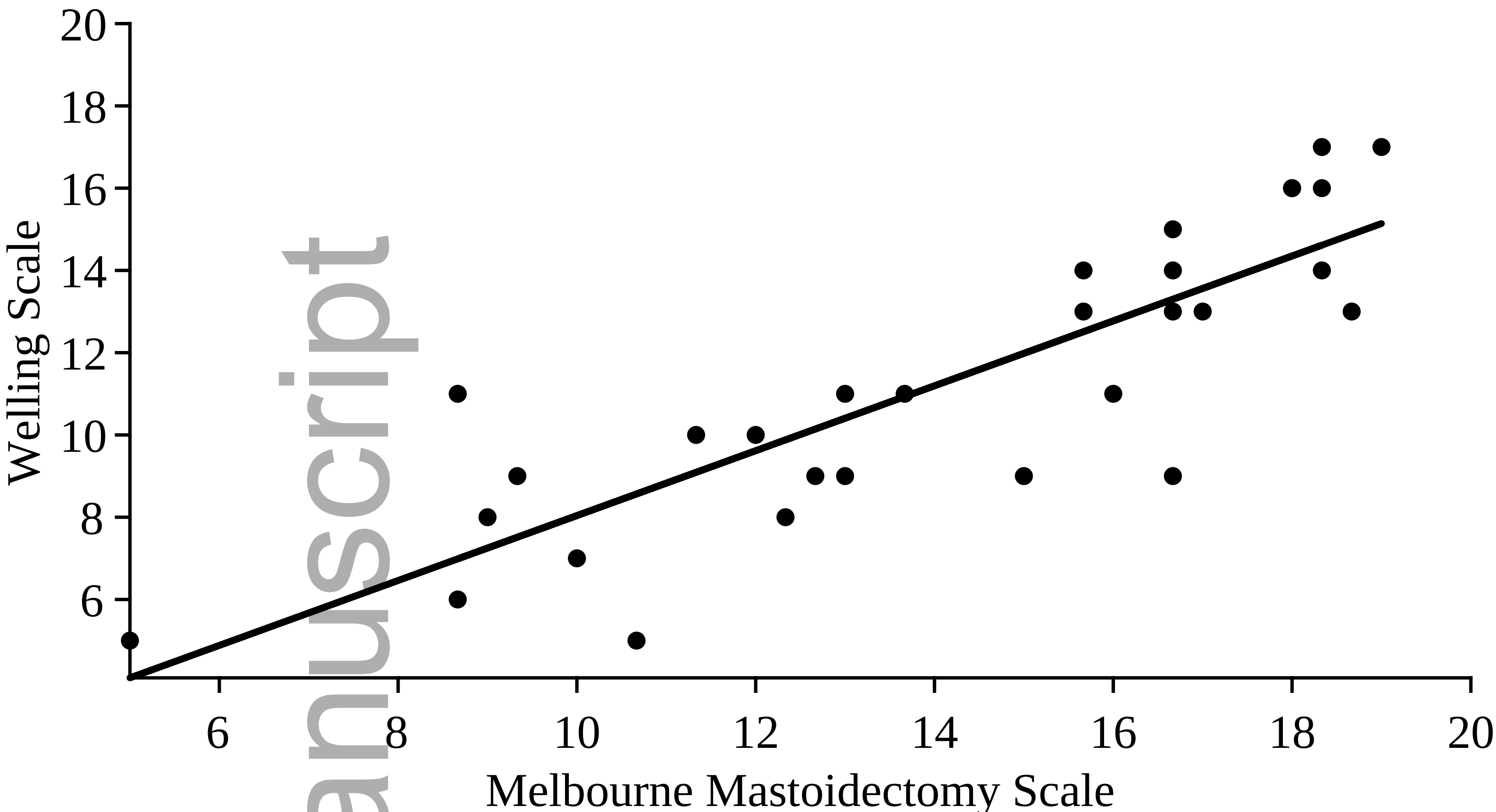
† These items represent major complications of the procedure and damage of the marked structures can class the dissection as unacceptable regardless of overall score.

‡ Superscripted numbers ([1-20]) represent the dependency of that item on a previous item on the scale denoted by the number.

coa_13569_f1.jpeg



coa_13569_f2.eps



coa_13569_f3.eps