

Concordance between independent first-line assessors in the Western Australian Audit of Surgical Mortality

Franca Itotoh, Natalie Zorbas, Andrew Lukman and R James Aitken 

Western Australian Audit of Surgical Mortality, Research, Audit and Academic Surgery Division, Royal Australasian College of Surgeons, Perth, Western Australia, Australia

Key words

audit, mortality, validation.

Correspondence

Mr R James. Aitken, Western Australian Audit of Surgical Mortality, Royal Australasian College of Surgeons, Suite 1, Ground Floor, 216 Stirling Highway, Claremont, WA 6010, Australia.
Email: rjaitken@me.com

F. Itotoh BDS, MSc; **N. Zorbas** BSocWk, MPH; **A. Lukman** BMed, GradDiplInfectDis; **R James Aitken** MS, FRACS.

Accepted for publication 11 July 2021.

doi: 10.1111/ans.17095

Abstract

Background: The Western Australian Audit of Surgical Mortality independently reviews deaths under a surgeon. A surgical case form completed by the surgeon is reviewed by a first-line assessor, who in 15% of cases recommends a detailed case note review. The aim of this study was to assess concordance between the original first-line assessment (FLA1) and an additional first-line assessment (FLA2) in relation to the necessity for a second-line assessment (SLA) and recognition of clinical management issues.

Methods: Experienced assessors ($n = 39$) representing different specialties were invited to be FLA2 assessors for 170 random cases. They completed identical FLA forms. FLA2 assessors were consultant surgeons with experience of the peer review process. Inter-rater reliability was determined using Gwet's Agreement Coefficient.

Results: A total of 170 FLAs, 158 (92.9%) were returned. Concordance between assessors for recommending an SLA was 70.5% (110 of 156), with an SLA recommended for 16 (10.3%) FLAs but not for 94 (60.3%) FLAs. When there was non-concordance, FLA1 assessors requested more SLAs than did FLA2 assessors (29 [18.6%] vs. 17 [10.9%], respectively). Three adverse events were recorded. In two the assessors differed, with one recording an adverse event and the other recording an area of concern. Additionally, in one instance the assessors differed on the preventability of the event but agreed that it may have contributed to death.

Conclusion: Concordance between assessors ranged between 'moderate' (requests for an SLA) through to 'almost perfect' (presence of an adverse event). This suggests the FLA screening process is reliable.

INTRODUCTION

The Western Australian Audit of Surgical Mortality (WAASM) is an external, independent, peer review audit of deaths that occur under the care of a surgeon. This includes where a patient was under the care of a surgeon, was under the care of another medical specialist and underwent a surgical procedure, or died within 30 days of a surgical procedure following discharge from hospital.

WAASM commenced in 2001. It was established nationally by the Royal Australasian College of Surgeons (RACS) in 2005 as the Australian and New Zealand Audit of Surgical Mortality (ANZASM).^{1,2} The methods have previously been described in detail.^{1,2} Patients admitted for palliative care who have no surgical procedure are excluded from the audit.³

In brief, WAASM is advised of deaths via the Western Australian (WA) Department of Health, hospitals and occasionally by treating surgeons. The treating surgeon is asked to complete a semi-structured surgical case form (SCF). This is reviewed by a first-line assessor—a consultant surgeon from the same specialty as the treating surgeon but working in a different hospital to where the death occurred.

The first-line assessment (FLA) is based on de-identified information in the SCF provided by the treating surgeon. The first-line assessor may finalise the review at this stage. The WAASM then sends a feedback letter to the treating surgeon and the case is completed. Alternatively, the first-line assessor may recommend that the case undergo a comprehensive second-line assessment (SLA), which involves an examination of the patient's medical records and

a detailed, written case note review (CNR). Between 2002 and 2018, 14.9% of FLAs were referred for SLA.⁴

The first-line assessors are gatekeepers. They must identify cases that do not require an SLA, but equally not miss any for which a more detailed analysis would be justified. Although ANZASM has undertaken other concordance studies, not all focused on concordance between FLAs.^{5–7} The only other ANZASM study that measured FLA concordance differed from the current study in both size (94 cases) and study design (FLA2 assessors were unaware they were completing a validation assessment).⁸

This study had two aims. The first was to ascertain consistency between two independent assessors when requesting an SLA. The second was to ascertain consistency between assessors in the identification and classification of clinical management issues (CMIs). These are categorised as being: an area for consideration (where the clinician believes areas of care could have been improved or done differently but recognises that this may be an area of debate), an area of concern (where the clinician believes that areas of care should have been better), or an adverse event (an unintended injury caused by medical management rather than by the disease process, which is sufficiently serious to lead to prolonged hospitalisation, or to temporary or permanent impairment or disability of the patient at the time of discharge, or which contributes to or causes death).

METHODS

This was a prospective observational study. Between 1 January 2019 and 30 June 2019, 341 original FLAs (FLA1s) were completed. For the purposes of this study, additional FLAs (FLA2) were undertaken in 170 (50%) randomly selected cases, by independent assessors. These assessors were unaware of the outcome of the FLA1s. The nine surgical specialties were represented in proportion to the total cases received from each specialty (Table 1). As there were no cases for Ophthalmology, Obstetrics and Gynaecology, and Oral/Maxillofacial Surgery within the period, these specialties were not included.

Table 1 Surgical specialties and FLA2 response rate

Specialties	Number of assessments distributed to FLA2 assessors	Number of assessments returned from FLA2 assessors (%)
General surgery	72	67 (93.1)
Orthopaedic surgery	34	34 (100)
Cardiothoracic surgery	20	20 (100)
Vascular surgery	14	11 (78.6)
Plastic surgery	12	8 (66.7)
Urology	10	10 (100)
Neurosurgery	4	4 (100)
Paediatric surgery	2	2 (100)
Otolaryngology, head and neck surgery	2	2 (100)
Total	170	158 (92.94)

The FLA2s were done for the purpose of this study only and did not alter the outcome of any case as part of the assessment process. Whilst the FLA1s were undertaken online via the Fellows Interface (the web-based application for electronically completing and submitting SCFs and FLAs), the FLA2s were completed on paper FLA forms provided to the FLA2 assessors and accompanied by de-identified paper copies of the SCFs.

The FLA2 assessors were sought from consultant surgeons who had between 1 to 18 years' experience in assessing cases and had been assigned a WAASM FLA between 1 January 2019 and 30 June 2019. Each assessor was assigned between one and five FLA2s according to the number of cases selected for each specialty.

Statistics

The inter-rater concordance for this study was determined using Gwet's Agreement Coefficient (AC1), which provides a stable inter-rater reliability coefficient not affected by prevalence and marginal probability. The null hypothesis of each test of significance (p value) is that the Gwet's AC1 score equals zero. The KAPPAETC module in Stata 15.1 (Stata Corporation, College Station, TX, USA) was used to calculate the Gwet's AC1 and percent agreement scores. The following interpretation of the scores, as suggested by Landis and Koch,⁹ was used: no agreement (<0), poor agreement (0.00–0.19), fair agreement (0.2–0.39), moderate agreement (0.40–0.59), substantial agreement (0.60–0.79) and almost perfect agreement (0.80–1.00).

RESULTS

The FLA2 assessors returned 158 of 170 (92.9%) FLAs.

Both the FLA1s and FLA2s had missing data, thus the denominator varies for different data points (Table 2). Missing data was greater for the FLA1s. This is likely because the FLA1 assessors did not fully complete the assessment once they had decided that the case should progress to an SLA.

In 110 of 156 (70.5%) FLAs, there was concordance between FLA1 and FLA2 assessors regarding the requirement or otherwise for an SLA (Gwet AC1 score 0.53, indicating moderate agreement). Both assessors agreed that 16 of 156 (10.3%) FLAs required an SLA and 94 of 156 (60.3%) FLAs did not. The FLA1 assessor, but not the FLA2 assessor, requested an SLA in 29 (18.6%) FLAs. Conversely, the FLA2 assessor, but not the FLA1 assessor, requested an SLA in 17 (10.9%) FLAs.

The difference by specialty between assessors requesting an SLA is shown in Table 3. The variations were more notable in the major specialties—Orthopaedic Surgery and General Surgery. Fewer FLA2 assessors recommended an SLA.

In 111 of 146 (76.0%) FLAs, there was concordance between both assessors regarding the presence or absence of a CMI (Gwet AC1 score 0.58, indicating moderate agreement). Both assessors identified CMIs in 27 of 146 (18.5%) FLAs and no CMIs in 84 of 146 (57.5%) FLAs (Fig. 1). Where there was concordance regarding the presence of a CMI, there was substantial to almost perfect agreement in the category of CMI (Gwet AC1 score: area for consideration, 0.72; area of concern, 0.82; and adverse event, 0.99).

Table 2 Concordance between FLA1 and FLA2 assessors

Variable	Number	Concordance (%)	Gwet's AC1 score	Missing data (%)
Should this case progress to SLA?	156	70.5	0.53	2/158 (1.3)
Were there any CMIs?	146	76.0	0.58	12/158 (7.6)
Death category	155	42.0	0.30	3/158 (1.9)
Enough information to reach a conclusion?	158	74.7	0.63	0/158 (0.0)
Should an operation have been performed?	141	79.4	0.74	17/158 (10.8)
Assessor's view of overall risk of death	149	53.0	0.44	9/158 (5.7)
Was the patient treated in a Critical Care Unit during this admission?	153	94.8	0.90	5/158 (3.2)
Should this patient have been provided critical care in the ICU?	47	100	1.00	0/47 (0.0)
Should this patient have been provided critical care in the HDU?	37	91.9	0.91	10/47 (21.3)
Was the decision on the use of DVT prophylaxis appropriate?	152	69.1	0.64	6/158 (3.8)
Was fluid balance an issue in this case?	154	57.1	0.46	4/158 (2.5)
Would it be beneficial for this case to undergo RCA?	149	75.2	0.71	9/158 (5.7)
If an operation occurred, could management have been improved in the following areas:				
Preoperative management/preparation	143	67.8	0.59	15/158 (9.5)
Decision to operate at all	147	76.2	0.70	11/158 (7.0)
Choice of operation	146	75.3	0.68	12/158 (7.6)
Timing of operation	145	77.2	0.70	13/158 (8.2)
Intraoperative/technical management of surgery	144	81.3	0.76	14/158 (8.9)
Grade/experience of surgeon deciding	145	80.7	0.77	13/158 (8.2)
Grade/experience of surgeon operating	144	81.2	0.77	14/158 (8.9)
Postoperative care	144	75.0	0.68	14/158 (8.9)

AC, agreement coefficient; CMIs, clinical management issues; DVT, deep vein thrombosis; HDU, high dependency unit; ICU, intensive care unit; RCA, root cause analysis; SLA, second-line assessment.

Table 3 Variance between FLA1 and FLA2 assessors for SLA recommendation by surgical speciality

	SLAs recommended by FLA1 assessors	SLAs recommended by FLA2 assessors
Cardiothoracic surgery	3	2
General surgery	11	7
Neurosurgery	1	0
Orthopaedic surgery	12	4
Plastic surgery	0	1
Urology	0	2
Vascular surgery	2	1
Total, n (%)	29 (18.6)	17 (10.9)

SLA, second-line assessment.

In 15 of 146 (10.3%) FLAs, the FLA1 assessor but not the FLA2 assessor indicated a CMI. Conversely, in 20 of 146 (13.7%) FLAs, the FLA2 assessor but not the FLA1 assessor indicated a CMI (Fig. 1).

For the 27 FLAs where both assessors agreed there was a CMI, both assessors classified the CMI in the same category in 16 (59.3%) FLAs. In six (22.2%) FLAs, the assessors differed in CMI categories assigned. For example, some assessors classified the CMI as an area for consideration and others as an area of concern, or some assessors classified the CMI as an area of concern and others as an adverse event. There was no instance where one assessor classified a CMI as an area for consideration and another

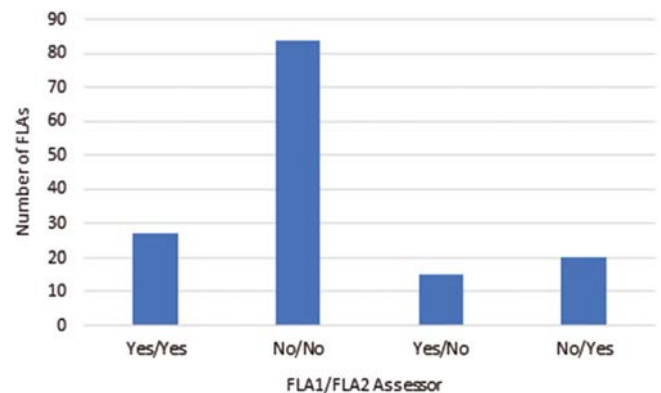


Fig 1. Concordance between FLA1 and FLA2 assessors regarding presence/absence of clinical management issues.

classified it as an adverse event. In the remaining five (18.5%) FLAs, some assessors classified the CMI as an area for consideration whilst the others did not categorise it, only indicating the presence of a CMI.

In 35 of 146 (24.0%) FLAs, assessors differed regarding the presence or absence of a CMI. Of these, 23 (65.7%) were instances where one assessor had stated that there was a CMI, categorised as an area for consideration, and another had recorded no CMI. These are minor variations/differences. In the remaining 12 (34.3%), one assessor had stated there was no CMI and another had stated there was a CMI categorised as an area of concern. These are more considerable variations/differences between assessors.

There were three cases where an adverse event was identified. In the first, the FLA2 assessor classified the CMI as an adverse event that caused death, whilst the FLA1 assessor identified it as an area of concern that may have contributed to death. Both assessors agreed that the CMI was probably preventable.

In the second case, both assessors agreed that the CMI was an adverse event that may have contributed to death. However, the FLA1 assessor saw this as probably preventable, whilst the FLA2 assessor considered that it was probably not preventable.

In the third case, the FLA1 assessor classified the CMI as an adverse event and the FLA2 assessor identified it as an area of concern. Both agreed that the CMI caused death and was probably preventable.

DISCUSSION

The first part of this study showed that cases requiring an SLA had one undertaken. The second part showed substantial to almost perfect concordance between assessors for serious CMIs. These data strongly suggest that while an FLA inevitably has a degree of subjectivity, it is robust, reproducible and endorses the role of the FLA assessor as a gatekeeper.

These observations suggest ANZASM's external, peer review process could be extended to surgical patients where death did not occur. It would not be practical, nor indeed necessary, to review every deviation from normal care. However, major deviations such as unplanned returns to theatre, re-admissions or sentinel complications, such as a pulmonary embolus, could undergo an external assessment to identify those requiring more detailed scrutiny. Such external scrutiny would be of particular value in smaller departments.

When considering the differences in requesting an SLA, it was because the FLA1 assessor rather than the FLA2 assessor sought an SLA. There are several likely reasons for this. The FLA2 assessors were drawn from an experienced pool of surgeons familiar with the peer review process and perhaps more confident in drawing conclusions from the FLA. By contrast, FLA1 assessors were selected from a wider pool, with some not having completed a peer review, being new to the audit's process. Completing an FLA can be difficult if the treating surgeon has not completed the SCF fully and, in cases of uncertainty, a less experienced FLA1 assessor would be more likely to recommend an SLA.

The FLA2 assessors were not blinded to the study, so they were aware they were not influencing the outcome of an actual WAASM process. Thus, their inclination to recommend an SLA may have been reduced. The clear and important finding is that greater tendency of the FLA1 assessors to recommend an SLA, indicates that cases requiring an SLA were unlikely to be overlooked.

A detailed SLA is time consuming for the WAASM team, hospital medical records staff and second-line assessors, as well as being an expensive undertaking (including postage and courier costs). It cannot be justified for the vast majority of deaths in which care was entirely appropriate. Conversely, it is important that first-line assessors do not miss cases that need an SLA. The role of first-line assessors in accurately recognising these cases is critical. It is likely that the number of SLAs could be reduced if treating surgeons were to complete the SCF with all the necessary information. The

ANZASM online electronic database prevents submission of the SCF until a significant proportion of the questions have been answered. However, it does not ensure the quality or extent of information provided in the free text questions.

Identification of CMIs is integral to WAASM's objectives by highlighting lessons that can be learned and the potential for change, not only in terms of individual surgeons and specific cases, but also in relation to systems and processes that impact on the care of surgical patients. Feedback and any recommendations arising from SLAs are disseminated through state and national publications. The WA Department of Health's Review of Death Policy requires that patient deaths are categorised in terms of preventability. This is done using the Health Roundtable criteria, which classifies deaths into five categories.¹⁰ Hospital clinical governance units are notified by individual surgeons of deaths identified as category 4 (preventable death where steps may not have been taken to prevent it) and category 5 (unexpected death resulting from medical intervention), as advised in WAASM feedback letters.

The observation in this study that there was a high degree of assessor concordance, aligns with previous ANZASM studies where concordance was also high in various aspects of the audit process.⁵⁻⁷

CONCLUSION

The medical profession has been granted the great privilege of self-regulation in the maintenance of clinical safety and quality. ANZASM is a significant safety and quality commitment by RACS and its Fellows. Previous ANZASM studies have shown that surgeons place great value on the peer review process (FLA and SLA).^{11,12} Any clinical review has some degree of subjectivity, so perfect agreement between assessors was not expected. However, it is reassuring that ANZASM's peer review process is robust and reproducible. Concordance could be further enhanced if the SCF were completed appropriately, thereby providing assessors with the detailed information required for thorough peer reviews.

ACKNOWLEDGEMENTS

The WAASM would like to acknowledge the WA Department of Health for funding the audit. Appreciation is extended to Mr. Dylan Hansen for statistical advice and support. We thank ANZASM management for review and oversight.

AUTHOR CONTRIBUTIONS

Franca Itotoh: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; supervision; validation; visualization; writing - original draft; writing-review & editing. **Natalie Zorbas:** Conceptualization; data curation; investigation; methodology; project administration; resources; validation; visualization; writing - original draft; writing-review & editing. **Andrew Lukman:** Resources; writing-review & editing. **Robert Aitken:** Conceptualization; methodology; supervision; writing - original draft; writing-review & editing.

Conflict of interest

None declared.

REFERENCES

1. Australian and New Zealand Audit of Surgical Mortality. Australian and New Zealand Audit of Surgical Mortality. https://umbraco.surgeons.org/media/3607/anzasm_2009_annual_report.pdf. Accessed 30 Sep 2020.
2. Western Australia Audit of Surgical Mortality. WAASM Reports. https://www.surgeons.org/-/media/Project/RACS/surgeons-org/files/surgical-mortality-audits/waasm/2020-10-14-RPT-WAASM_Annual_Report_2020.pdf?rev=c9774c82c34e4886849d0f6e1560d404&hash=EF2683160A1EE550671A16A2D323B6A5. Accessed 20 Oct 2020.
3. Raju RS, Guy GS, Majid AJ, Babidge W, Maddern GJ. The Australian and New Zealand Audit of Surgical Mortality – birth, deaths, and carriage. *Ann Surg*. 2015;**261**:304–8. <https://doi.org/10.1097/SLA.581>.
4. Western Australian Audit of Surgical Mortality. WAASM Reports. <https://umbraco.surgeons.org/media/4438/2019-10-08-rpt-waasm-2019-report.pdf>. Accessed 20 Mar 2020.
5. Raju RS, Guy GS, Field JBF, Kiroff GK, Babidge W, Maddern GY. Australian and New Zealand Audit of Surgical Mortality: concordance between reported and audited clinical events and delays in management in surgical mortality patients. *ANZ J Surg*. 2014;**84**:618–23.
6. Rey-Conde T, Shakya R, Allen J, Clarke E, North JB. Surgical mortality audit data validity. *ANZ J Surg*. 2016;**86**:644–7.
7. Hansen D, Hansen E, Retegan C, Morphet J, Beiles CB. Validation of data submitted by the treating surgeon in the Victorian Audit of Surgical Mortality. *ANZ J Surg*. 2019;**89**:16–9.
8. Victorian Audit of Surgical Mortality (VASM). First-Line Assessment Validation Audit 2017 Report. <https://www.surgeons.org/research-audit/surgical-mortality-audits/regional-audits/vasm/reports-publications>. Accessed 20 Mar 2020.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;**33**:159–74.
10. Review of Death Policy. Health Roundtable criteria for preventability of death. Government of Western Australia, Department of Health. <https://ww2.health.wa.gov.au/-/media/Files/Corporate/Policy-Frameworks/Clinical-Governance-Safety-and-Quality/Policy/Review-of-death/Review-of-death-policy.pdf>. Accessed 21 Jan 2021.
11. Semmens JB, Mountain JA, Sanfilipo FM, Barraclough JY, McKenzie A, Mukhtar SA, et al. Providers and consumers support the Western Australian audit of surgical mortality. *ANZ J Surg*. 2006;**76**:442–7.
12. Lui C, Boyle FM, Wysocki AP, Baker P, D'Souza A, Faint S, et al. How participation in surgical mortality audit impacts surgical practice. *BMC Surg*. 2017;**17**:42.