ORIGINAL ARTICLE

ANZJSurg.com

Validation of data submitted by the treating surgeon in the Victorian Audit of Surgical Mortality

Dylan Hansen,* Emma Hansen,† Claudia Retegan,* Julia Morphet† and Charles Barry Beiles*

*Victorian Audit of Surgical Mortality (VASM), Royal Australasian College of Surgeons, Melbourne, Victoria, Australia and
†Nursing and Midwifery, Monash University, Melbourne, Victoria, Australia

Key words
clinical audit, healthcare quality assurance, mortality, surgery, validity.

Correspondence
Mr Charles Barry Beiles, Victorian Audit of Surgical Mortality (VASM), Royal Australasian College of Surgeons, 3/30-32 Fisher Street, Malvern East, Melbourne, VIC 3145, Australia.
Email: barryb2@optusnet.com.au

D. Hansen BSc (PharmSc); **E. Hansen** BNurs, GDip, AdvNurs (Emerg); **C. Retegan** BA, DipSc; **J. Morphet** PhD, MN (Ed); **C. B. Beiles** FRACS (Vasc).

Accepted for publication 15 September 2018.

doi: 10.1111/ans.14910

Abstract

Background: The Victorian Audit of Surgical Mortality (VASM) seeks to peer review all deaths associated with surgical care in Victoria, Australia. The effectiveness of the VASM as an educational and quality improvement tool is dependent on the accuracy of source data it receives. We aimed to examine the accuracy and quality of source data provided by the treating surgeon for peer review, and the inter-rater concordance level between the external validator findings and the treating surgeon.

Methods: Of the 629 cases that completed the VASM audit second-line peer review process over a 4-year period (from 1 July 2012 to 30 June 2016), a total of 32 (5%) were randomly selected for the external validation process. The blinded external validator was impartial to the VASM audit, and was provided only de-identified patient medical records. The analysis for the checked and validated data points and their concordance was determined using Gwet's agreement coefficient, which provides a stable inter-rater reliability coefficient not affected by prevalence and marginal probability.

Results: The inter-rater concordance analysis suggested that there is a high level of agreement (82.9% overall) between the treating surgeon and external validator. The use of thromboembolism deterrent stockings was the only variable where agreement was poor (52.4%) with a Gwet score of 0.10 (−0.40 to 0.60).

Conclusion: The inter-rater concordance analysis results support the validity of the VASM process, which is dependent on the accuracy of data submitted by the treating surgeon.

Introduction

The Victorian Audit of Surgical Mortality (VASM) is part of the Australian and New Zealand Audit of Surgical Mortality (ANZASM), which is a network of regionally based audits of surgical mortality.[1] These audits peer review all deaths associated with surgical care to improve the quality of health care through feedback and education by monitoring and reporting on patient management issues that may occur during surgical admissions.[2–5]

In order to ensure that the peer review process remains a reliable assessment, it is essential to verify the accuracy of the information received by the audit office from the treating surgeon.[6–8] Source documentation forms the basis of the assessment and reporting process, as all data submitted by the treating surgeon are subjected to a first-line assessment. The first-line assessment is completed by an independent surgeon from the same specialty who decides whether to refer the case for a second-line assessment (SLA). A previous verification process was performed in Queensland[9] which suggested that a wider study be performed to confirm the level of concordance between the treating surgeon and the SLA. This is the first follow-up to the Queensland study which made the recommendation to verify source documentation of another region's surgical mortality data.

Methods

This was a retrospective observational study and has enhanced the methodology used in Queensland by blinding the validator to the responses from the treating surgeon and applying a more robust statistical tool for inter-rater analysis. This verification was necessary to build and maintain confidence in the findings that the VASM presents in driving health policies.

A total of 629 cases in the VASM were identified over a 4-year period (between 1 July 2012 and 30 June 2016) that had been

referred for and completed an SLA. From these, a total of 32 (5%) were randomly selected for external validation using a certified copy of hospital patient medical records. The required sample size of 32 was determined by using the guidelines outlined by Sim and Wright.[10] We used their recommendation of a null hypothesis value of our kappa like statistic to be 0.00 with a desired power of 80% to detect a statistically significant *P*-value of ≤0.05 using a two-tailed test.

The certified copy of the medical records, supplied by the hospital, permits replication and verification of the accuracy of the source documents submitted by the treating surgeon. This process permitted the external auditor to have access to the same information as the treating surgeon and reproduce the same collection of data points. This allowed for monitoring of compliance and verification of the integrity of the VASM audit data. All audit reports were retained and stored securely for verification.

The external validator, a healthcare professional not affiliated with the VASM, was provided only de-identified patient medical records in order to blind them to the identity of the hospital, surgeon and patient. They were also blinded to the data points submitted by the treating surgeon, allowing the reproduction and evaluation of the data points required by the audit process. Analysis of the data was performed by a separate researcher.

The study cohort was selected using proportionate stratified sampling. The 32 cases were randomly selected to be proportional to the population of 629 audited cases representing the strata year of death and the nine different specialties (cardiothoracic surgery, general surgery, gynaecology, neurosurgery, orthopaedic surgery, paediatric surgery, plastic surgery, urology and vascular surgery). The data points selected for checks from the source documents were clinical, tangible and verifiable (patient demographics, co-morbidity profile, American Society of Anesthesiologists (ASA) status, diagnosis, cause of death, transfers, critical care unit care, operation details, post-operative complications, return to theatre, infection profile and trauma involvement), rather than subjective or reflective data points such as alternative clinical decision for operative methods. Otolaryngology was not selected due to very low numbers of cases available for validation.

Statistical analysis

The independent validation data were compared with the source data provided by the treating surgeon contained within the surgical case form during the audit process. For the 26 dichotomous and eight nominal variables, Gwet's Agreement Coefficient (AC1) with 95% confidence intervals and *P*-values were calculated. Gwet's AC1 scores were chosen over Cohen's kappa as they are less affected by prevalence and marginal probability.[11,12] An agreement percentage was used for the nominal variables. The null hypothesis of each *P*-value test is that the Gwet's AC1 score is equal to zero.

The KAPPAETC module in Stata 15.0 (StataCorp, College Station, TX, USA) was used to calculate Gwet's AC1 and percent agreement scores. The intervals suggested by Landis and Koch[13] provided in Table 1 were used.

Results

The overall agreement for dichotomous and nominal variables was 82.9% with a Gwet AC score of 0.82 (0.79–0.84) indicating almost perfect agreement between the original treating surgeon and the external independent blinded validator. Perfect concordance was found in two of the dichotomous variables, gender and operative status, while there was perfect concordance in just one of the nominal variables, surgical specialty of the consultant in charge of the patient (Tables 2,3).

Poor agreement, as defined by a Gwet score of 0–0.19, was observed in only one variable, the use of thromboembolic deterrent (TED) stockings with a concordance of 52.4% and agreement of 0.10 (−0.40 to 0.60). For this variable, in five cases, the treating surgeon identified that TED stockings were used when the external validator did not. Conversely, there were seven cases where the external validator identified the use of TED stockings when the treating surgeon did not, including two instances where the treating surgeon indicated that no forms of deep venous thrombosis prophylaxis were used. Fair agreement was observed in three more variables: those were presence of post-operative complications 0.21 (−0.16 to 0.59), use of heparin 0.30 (−0.19 to 0.79) and timing of the operation 0.36 (0.11 to 0.61).

The fields with the most missing data for verification were patient status, operation abandoned on finding a terminal situation, timing of operation and date of birth (Table 2).

Discussion

The overall agreement between the data provided to VASM on the surgical case form and the hospital medical records was high, as outlined in Tables 2 and 3. Accurate information is crucial to obtain meaningful outcomes when surgical peer review is performed. The high concordance achieved for these data points compares favourably with national and international figures[8,9] which reassure the audit office that the peer review process is based on accurate source information provided by the treating surgeon. This is similar to other source validations conducted on other Australian clinical audits such as our Queensland counterparts who found an overall agreement of 98.2%[9] or the Victorian Cardiac Outcomes Registry, which found a 97.4% agreement between their data and hospital medical records.[14]

It is noted that the use of TED stockings appears in the nursing notes rather than in the drug charts, as the task of prescribing deep venous thrombosis prophylaxis often falls to junior members of a clinical team[15] and the responsibility of completing the VASM

Table 1 Landis and Koch agreement level

Gwet's AC1 score	Agreement
<0	No agreement
0.00–0.19	Poor agreement
0.20–0.39	Fair agreement
0.40–0.59	Moderate agreement
0.60–0.79	Substantial agreement
0.80–1.00	Almost perfect agreement

AC, agreement coefficient.

Table 2 Gwet's AC1 for dichotomous and nominal variables between the treating surgeon and validator

Item	*n*	Concordance (%)	Gwet's AC1 score	95% CI	*P*-value	Missing data (*n*/d)
Gender	31	100	1.00	1.00 to 1.00	—	1/32 (3.1%)
Specialty of consultant surgeon in charge of patient	31	100	1.00	1.00 to 1.00	—	1/32 (3.1%)
Admission type	32	87.5	0.79	0.57 to 1.00	<0.0001	0/32 (0.0%)
Patient status†	20	80.0	0.74	0.48 to 1.00	<0.0001	12/32 (37.5%)
Were there significant coexisting factors increasing risk of death	30	83.3	0.79	0.58 to 1.00	<0.0001	2/32 (6.3%)
Cofactor – advanced malignancy	24	87.5	0.85	0.65 to 1.00	<0.0001	0/24 (0.0%)
Cofactor – age	24	79.2	0.58	0.23 to 0.93	0.0022	0/24 (0.0%)
Cofactor – cardiovascular	24	70.8	0.46	0.07 to 0.86	0.0242	0/24 (0.0%)
Cofactor – diabetes	24	87.5	0.82	0.59 to 1.00	<0.0001	0/24 (0.0%)
Cofactor – hepatic	24	91.7	0.90	0.75 to 1.00	<0.0001	0/24 (0.0%)
Cofactor – neurological	24	83.3	0.75	0.48 to 1.00	<0.0001	0/24 (0.0%)
Cofactor – obesity	24	87.5	0.85	0.65 to 1.00	<0.0001	0/24 (0.0%)
Cofactor – other	24	66.7	0.50	0.12 to 0.89	0.0133	0/24 (0.0%)
Cofactor – renal	24	83.3	0.72	0.42 to 1.00	0.0001	0/24 (0.0%)
Cofactor – respiratory	24	70.8	0.52	0.13 to 0.90	0.0104	0/24 (0.0%)
ASA grade	23	69.6	0.61	0.34 to 0.87	0.0001	9/32 (28.1%)
Was the patient transferred preoperatively	32	78.1	0.64	0.36 to 0.92	0.0001	0/32 (0.0%)
Was this patient treated in a critical care unit (ICU or HDU) during this admission	32	92.8	0.90	0.75 to 1.00	<0.0001	0/32 (0.0%)
Was an operation performed during the last admission	32	100	1.00	1.00 to 1.00	—	0/32 (0.0%)
Timing of operation	31	51.6	0.36	0.11 to 0.61	0.0061	11/42 (26.2%)
Was the operation abandoned on finding a terminal situation	30	90.0	0.88	0.75 to 1.00	<0.0001	12/42 (28.6%)
Was there a definable post-operative complication	30	60.0	0.21	−0.16 to 0.59	0.2574	1/31 (3.2%)
Was DVT prophylaxis used during the admission	31	87.1	0.79	0.57 to 1.00	<0.0001	1/32 (3.1%)
DVT prophylaxis – aspirin	21	85.7	0.82	0.58 to 1.00	<0.0001	0/21 (0.0%)
DVT prophylaxis – heparin	21	57.1	0.30	−0.19 to 0.79	0.2162	0/21 (0.0%)
DVT prophylaxis – other	21	90.5	0.90	0.73 to 1.00	<0.0001	0/21 (0.0%)
DVT prophylaxis – SCD	21	81.0	0.64	0.28 to 1.00	0.0014	0/21 (0.0%)
DVT prophylaxis – TED stocking	21	52.4	0.10	−0.40 to 0.60	0.6843	0/21 (0.0%)
DVT prophylaxis – warfarin	21	90.5	0.90	0.73 to 1.00	<0.0001	0/21 (0.0%)
Was there an unplanned return to theatre	31	87.1	0.78	0.55 to 1.00	<0.0001	1/31 (3.2%)
Was there an unplanned admission to a critical care unit	32	90.6	0.83	0.63 to 1.00	<0.0001	0/32 (0.0%)
Was there an unplanned readmission within 30 days of surgery	31	96.8	0.96	0.89 to 1.00	<0.0001	1/32 (3.1%)
Did this patient die with a clinically significant infection	31	74.2	0.52	0.20 to 0.85	0.0026	1/32 (3.1%)
Was trauma involved	31	100	1.00	1.00 to 1.00	—	1/32 (3.1%)

†Patient status: public or private. AC, agreement coefficient; ASA, American Society of Anesthesiologists; CI, confidence interval; DVT, deep venous thrombosis; HDU, high dependency unit; ICU, intensive care unit; *n*, number; *n*/d, number of observations unavailable/total observations possible (percentage observations unavailable of total observations possible); SCD, sequential compression device; TED, thromboembolism deterrent.

Table 3 Agreement between dates and free text variables

Item	*n*	Concordance (%)	95% CI	Missing data (*n*/d)
Date of birth	21	95.2	85.3–100	11/32 (34.4%)
Admission date	31	80.0	65.9–95.4	1/32 (3.1%)
Date of death	31	96.8	90.2–100	1/32 (3.1%)
Operation date/time	32	71.9	55.4–88.3	10/42 (23.8%)
Main surgical diagnosis on admission	32	78.1	63–93.3	0/32 (0.0%)
Confirmed main surgical diagnosis	32	75.0	59.1–90.9	0/32 (0.0%)
Final cause of death	30	70.0	52.6–87.4	2/32 (6.3%)
Description of operation	42	95.2	88.5–100	0/42 (0.0%)

CI, confidence interval; *n*, number; *n*/d, number of observations unavailable/total observations possible (percentage observations unavailable of total observations possible).

audit falls upon the consultant surgeon. This may account for the instances where the external validator identified the use of TED stockings but the treating surgeon did not.

Timing of operation concordance was 51.6% with fair agreement of 0.36 (0.11–0.61). Differences can be due to inaccurate recording of the preparation timeframe prior to surgery and post-surgery recovery timeframe. Russ *et al.*[16] and Putnam *et al.*[17] had similar findings for sign-in and sign-out records, particularly omissions in 39% of cases, largely due to uncertainty about when to record these precisely.[16,17] Adherence to surgical safety checklists and accurate completion of forms remain a challenge within hospital settings.[17]

Post-operative complications also had a fair agreement level of 0.21 (−0.16 to 0.59) with a concordance at 60% where complications had been identified as reported by the treating surgeon; however, additional complications may have been identified in the

nursing notes by the external validator which might explain this difference.

We have made two adjustments to the study design utilized by our Queensland counterparts. Our study blinded the external validator to the answers provided by the treating surgeon, thus removing a potential source of bias from the study design.[18] Gwet's AC1 score calculation was used, which accounts for the probability that the treating surgeon and the external validator may agree based on chance alone.[11] These improvements to study design and analysis may explain why the results are slightly lower than our Queensland counterparts.

Limitations

Due to the retrospective nature of the audit, both the VASM source file submission and the peer review assessments may involve some degree of subjectivity.

In the presence of conflicting information in the patient medical records, the external validator was unable to verify or resolve the conflict. A conflict could occur when two clinicians call the same condition two different things, for example sprain versus fracture, or the surgeon provides an ASA grade different to the anaesthetist. The audit team can verify additional clinical evidence if presented as supporting documentation supplied by the treating surgeon, which can be checked against the hospital medical records.

The external validator could potentially identify the originating hospital due to the format and style of the patient medical records. However, the validator had no access to any identifiers and was fully blinded to the source data provided by the treating surgeon to the audit office.

Reliance on clinical data reported by the treating surgeon can produce incomplete data collection and therefore missing data was handled by deletion. This potential exclusion bias is a limitation of this study although this only involved a small number of cases.

Conclusion

Inter-rater concordance analysis results support the validity of the audit process with high inter-rater agreements between the treating surgeon and the external validator. This reassures us that accurate clinical level information is provided to the audit office by the treating surgeons. This is necessary for the VASM to be an effective educational and quality improvement tool.

Conflicts of interest

None declared.

References

1. 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.
2. Russell EM, Bruce J, Krukowski ZH. Systematic review of the quality of surgical mortality monitoring. *Br. J. Surg.* 2003; **90**: 527–32.
3. Semmens JB, Aitken RJ, Sanfilippo FM, Mukhtar SA, Haynes NS, Mountain JA. The Western Australian Audit of Surgical Mortality: advancing surgical accountability. *Med. J. Aust.* 2005; **183**: 504–8.
4. Watters DA, Green AJ, van Rij A. Guidelines for surgical audit in Australia and New Zealand. *ANZ J. Surg.* 2006; **76**: 78–83.
5. Retegan C, Russell C, Harris D, Andrianopoulos N, Beiles CB. Evaluating the value and impact of the Victorian Audit of Surgical Mortality. *ANZ J. Surg.* 2013; **83**: 724–8.
6. Hansen D, Retegan C, Woodford N, Vinluan J, Beiles CB. Comparison of the Victorian Audit of Surgical Mortality with coronial cause of death. *ANZ J. Surg.* 2016; **86**: 437–41.
7. Magee MJ, Wright CD, McDonald D, Fernandez FG, Kozower BD. External validation of the Society of Thoracic Surgeons General Thoracic Surgery Database. *Ann. Thorac. Surg.* 2013; **96**: 1734–9.
8. Varmdal T, Ellekjaer H, Fjaertoft H, Indredavik B, Lydersen S, Bonaa KH. Inter-rater reliability of a national acute stroke register. *BMC. Res. Notes* 2015; **8**: 584.
9. Rey-Conde T, Shakya R, Allen J. Surgical mortality audit data validity. *ANZ J. Surg.* 2016; **86**: 644–7.
10. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys. Ther.* 2005; **85**: 257–68.
11. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med. Res. Methodol.* 2013; **13**: 61.
12. Walsh P, Thornton J, Asato J *et al*. Approaches to describing inter-rater reliability of the overall clinical appearance of febrile infants and toddlers in the emergency department. *PeerJ.* 2014; **2**: e651.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74.
14. Lefkovits J, Brennan A, Dinh D, Brien R, Driscoll A, Reid C. The Victorian Cardiac Outcomes Registry Annual Report 2015. Monash University, June 2016, Report No: 3; 74 pp. [Cited 6 Sept.] Available from URL: https://vcor.org.au/sites/default/files/VCOR_files/2015_vcor_annual_report.pdf
15. Noel SE, Millar JA. Current state of medical thromboprophylaxis in Australia. *Australas. Med. J.* 2014; **7**: 58–63.
16. Russ S, Rout S, Caris J *et al*. Measuring variation in use of the WHO surgical safety checklist in the operating room: a multicenter prospective cross-sectional study. *J. Am. Coll. Surg.* 2015; **220**: 1–11.e4.
17. Putnam LR, Levy SM, Sajid M *et al*. Multifaceted interventions improve adherence to the surgical checklist. *Surgery* 2014; **156**: 336–44.
18. Day SJ, Altman DG. Statistics notes: blinding in clinical trials and other studies. *BMJ* 2000; **321**: 504.