Machine learning to predict myocardial injury and death after non-cardiac surgery

Citation for published version:
Nolde, JM, Schlaich, MP, Sessler, DI, Mian, A, Corcoran, TB, Chow, CK, Chan, MTV, Borges, FK, McGillion, MH, Myles, PS, Mills, NL, Devereaux, PJ & Hillis, GS 2023, 'Machine learning to predict myocardial injury and death after non-cardiac surgery', *Anaesthesia: Peri-operative medicine, critical care and pain*. https://doi.org/10.1111/anae.16024

Digital Object Identifier (DOI):
10.1111/anae.16024

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Anaesthesia: Peri-operative medicine, critical care and pain

Original Article

Machine learning to predict myocardial injury and death after non-cardiac surgery

J. M. Nolde,[1] **M. P. Schlaich,**[2] **D. I. Sessler,**[3] **A. Mian,**[4] **T. B. Corcoran,**[5] **C. K. Chow,**[6] **M. T. V. Chan,**[7] **F. K. Borges,**[8] **M. H. McGillion,**[9] **P. S. Myles,**[10] **N. L. Mills,**[11] **P. J. Devereaux**[12] and **G. S. Hillis**[13]

1 PhD Student, 2 Professor, Dobney Hypertension Centre, Royal Perth Hospital Research Foundation, Perth, Australia
3 Professor and Chair, Department of Outcomes Research, Cleveland Clinic, Cleveland, OH, USA
4 Professor, School of Computer Science and Software Engineering, University of Western Australia, Perth, Australia
5 Clinical Professor and Director of Research, Department of Anaesthesia and Pain Medicine, Royal Perth Hospital, Perth, Australia and Medical School, University of Western Australia and Department of Anaesthesiology and Peri-operative Medicine, Alfred Hospital and Monash University, Melbourne, Australia
6 Professor and Academic Director, Westmead Applied Research Centre, Faculty of Medicine and Health, University of Sydney, Sydney, Australia and Department of Cardiology, Westmead Hospital, Sydney, Australia
7 Professor, Department of Anaesthesia and Intensive Care, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China
8 Assistant Professor, 9 Associate Professor, 12 Professor, McMaster University, Faculty of Health Sciences, Hamilton, ON, Canada and Population Health Research Institute, Hamilton, ON, Canada
10 Professor, Department of Anaesthesiology and Peri-operative Medicine, Alfred Hospital and Monash University, Melbourne, Australia
11 Professor, British Heart Foundation Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, UK and Usher Institute, University of Edinburgh, Edinburgh, UK
13 Clinical Professor, Medical School, University of Western Australia and Department of Cardiology, Royal Perth Hospital, Perth, Australia

Summary

Myocardial injury due to ischaemia within 30 days of non-cardiac surgery is prognostically relevant. We aimed to determine the discrimination, calibration, accuracy, sensitivity and specificity of single-layer and multiple-layer neural networks for myocardial injury and death within 30 postoperative days. We analysed data from 24,589 participants in the Vascular Events in Non-cardiac Surgery Patients Cohort Evaluation study. Validation was performed on a randomly selected subset of the study population. Discrimination for myocardial injury by single-layer vs. multiple-layer models generated areas (95%CI) under the receiver operating characteristic curve of: 0.70 (0.69–0.72) vs. 0.71 (0.70–0.73) with variables available before surgical referral, $p < 0.001$; 0.73 (0.72–0.75) vs. 0.75 (0.74–0.76) with additional variables available on admission, but before surgery, $p < 0.001$; and 0.76 (0.75–0.77) vs. 0.77 (0.76–0.78) with the addition of subsequent variables, $p < 0.001$. Discrimination for death by single-layer vs. multiple-layer models generated areas (95%CI) under the receiver operating characteristic curve of: 0.71 (0.66–0.76) vs. 0.74 (0.71–0.77) with variables available before surgical referral, $p = 0.04$; 0.78 (0.73–0.82) vs. 0.83 (0.79–0.86) with additional variables available on admission but before surgery, $p = 0.01$; and 0.87 (0.83–0.89) vs. 0.87 (0.85–0.90) with the addition of subsequent variables, $p = 0.52$. The accuracy of the multiple-layer model for myocardial injury and death with all variables was 70% and 89%, respectively.

Correspondence to: G. S. Hillis
Email: graham.hillis@health.wa.gov.au
Accepted: 17 March 2023

Keywords: machine learning; myocardial injury; non-cardiac surgery
Twitter: @markus_schlaich; @OutcomesRC; @AjmalSaeedMian; @clara_chow; @HighSTEACS

Introduction

Over 200 million patients have major non-cardiac surgery each year, around 10 million of whom have a major cardiovascular event within 30 days [1, 2]. These patients have increased short-term mortality, ill-health and decreased survival [2, 3].

Postoperative myocardial infarction and injury are the most frequent cardiovascular complications after non-cardiac surgery but are often undetected because symptoms are masked, electrocardiograms may not be routinely obtained and ischaemic changes may be subtle or missed [3, 4]. If routinely checked, a high-sensitivity troponin assay is elevated in approximately one in five patients during the first three postoperative days and, in 80–90% of cases, this is due to ischaemia [4]. Regardless of the mechanism, myocardial injury is associated with mortality and commonly does not fulfil the definition of myocardial infarction [4–8]. This has led to the concept of prognostically relevant myocardial injury after non-cardiac surgery (MINS) [9]. This can be easily, accurately, quantifiably and reproducibly measured and some of the harm associated with MINS is due to recurrent atherothrombotic events that can be reduced [6]. In addition, peri-operative haemodynamic and other physiological changes may be important contributors to MINS that could be corrected by earlier intervention.

Prediction of postoperative death and MINS is complex. Both are associated with several factors, including the underlying condition requiring surgery, age, health, fitness and peri-operative complications. Thus, scores associating these factors with peri-operative cardiovascular events have well-documented limitations and there is currently no clinically applied scoring system to predict MINS [2].

Machine (computer) learning associates variables and outcomes without explicit programming. It is well suited to situations where large amounts of data are acquired and where causal links are unclear or complex. Importantly, unlike traditional risk scores, these techniques are inherently adaptable, such that algorithms derived in one dataset can continually retrain and refine their accuracy in other populations and settings. Machine learning algorithms can also be integrated into modern physiological monitoring systems, providing clinicians with an automated, real-time estimate of risk. While a wide range of machine learning techniques are available, we selected neural networks because of their intrinsic ability to exploit complex relationships in data, including high-dimensional non-linear relationships and interactions [10].

We therefore tested the hypothesis that neural networks of varying complexity could predict death and MINS at 30 postoperative days, using data collected in the Vascular Events in Non-cardiac Surgery Patients Cohort Evaluation (VISION) study [4, 5, 9].

Methods

The VISION study was approved by the institutional review board at each participating site and has been reported in detail [4, 5, 9]. In brief, we recruited patients aged ≥ 45 y who had non-cardiac surgery and then stayed at least one night in hospital. This analysis is of 24,589 patients who had postoperative troponin T measured with a high-sensitivity assay [4]. Table 1 lists the peri-operative variables that we analysed for their associations with MINS and death. We defined a history of coronary artery disease as any of: angina; myocardial infarction or acute coronary syndrome; a segmental cardiac wall motion abnormality on echocardiography or a segmental fixed defect on radionuclide imaging; a positive radionuclide exercise, echocardiographic exercise or pharmacological cardiovascular stress test demonstrating cardiac ischaemia; coronary angiographic or CT coronary angiographic evidence of atherosclerotic stenosis ≥ 50% of the diameter of any coronary artery; or pathological Q waves in two contiguous ECG leads. We defined high-risk coronary artery disease as a diagnosis ≤ 6 months before non-cardiac surgery of myocardial infarction or acute coronary syndrome, or Canadian Cardiovascular Society Class 3 or 4 angina. We substituted an adjacent measurement for missing values for some variables, for instance, pre-operative blood pressure for unrecorded intra-operative blood pressure. We ran K-Nearest neighbour imputations for missing data [11].

We analysed high-sensitivity troponin T (hs-TnT) in blood sampled 6–12 h and 1, 2 and 3 days after surgery and asked centres to obtain hs-TnT measurements and perform serial electrocardiograms if a participant experienced an ischaemic symptom. We assessed patients for features consistent with

Table 1 Variables analysed by the single-layer and multiple-layer neural networks in three sequential groups, categorised by when these usually become available. The discriminatory contributions of 12 clusters of variables (bold) are detailed in Fig. 1.

First	Second	Third
Demographic	**Clinical and blood results**	**Intra-operative measurements**
Age	Cancer present	SBP < 100 mmHg or > 160 mmHg
Sex	Surgery for cancer	Heart rate < 55 bpm
Weight	Metastatic cancer	Duration of heart rate < 55 bpm
Height	Fracture present	
Ethnicity	Chronic pain	**Recovery room measurements**
	Dialysis	SBP < 100 mmHg or > 160 mmHg
Care status	NT-proBNP	Heart rate < 55 bpm
Nursing home	Haemoglobin	Duration of heart rate < 55 bpm
ADL assistance	Creatinine	
Bed $h.day^{-1}$		**Postoperative days 1–4**
		SBP < 100 mmHg or > 160 mmHg
Smoking history	**Admission measurements**	Heart rate < 55 bpm
	SBP	Duration of heart rate < 55 bpm
Medical conditions	DBP	
Atrial fibrillation	Heart rate	**Postoperative ventilation and oxygenation**
Heart failure		
Coronary artery disease	**Type of surgery**	S_pO_2 level if < 90%
High-risk coronary artery disease		Duration of S_pO_2 < 90%
Cardiac arrest		Respiratory rate if < $10.min^{-1}$
Aortic stenosis	**Type of anaesthesia**	Duration of respiratory rate < $10.min^{-1}$
DVT or PE		Supplemental oxygen
Stroke		Ventilatory support
Sleep apnoea		Naloxone given
Peripheral arterial disease		
Hypertension		
COPD		
Diabetes		
Number of pregnancies		

ADL, activities of daily living; DVT, deep vein thrombosis; PE, pulmonary embolism; COPD, chronic obstructive lung disease; NT-proBNP, N-terminal pro B-type natriuretic peptide; SBP, systolic blood pressure; DBP, diastolic blood pressure; S_pO_2, pulse oximetry.

myocardial ischaemia if a hs-TnT level ≥ 14 $ng.l^{-1}$ was recorded [4]. In three centres, ischaemic features were identified retrospectively from clinical notes as hs-TnT levels were unavailable on site. Experts adjudicated from clinical records whether participants with hs-TnT ≥ 14 $ng.l^{-1}$ exhibited evidence of myocardial ischaemia or a clear non-ischaemic mechanism during or after surgery. We telephoned participants or their families 30 days after surgery to document whether they had died.

The primary outcome was MINS within 30 days of surgery, defined as postoperative hs-TnT ≥ 65 $ng.l^{-1}$ or between 20 and 64 $ng.l^{-1}$ with an absolute change of ≥ 5 $ng.l^{-1}$ between two samples, judged to be due to myocardial ischaemia, which was assumed unless there was clear evidence of a non-ischaemic mechanism [4]. The methodology used to determine these thresholds has been previously reported [4]. The secondary outcome was death at 30 days.

We used Keras and Tensorflow for Python 3 to create neural networks [12, 13]. Class weights were implemented for the training process to reflect the imbalance between event and non-event cases (online Supporting Information Appendix S1). We used a leave-one-out-bootstrap method with 250 repetitions [14, 15]. This process used the proportion of the dataset not included in each bootstrapped sample (so-called out-of-bag sample, which is about 38% of the original sample with some random variability in each iteration), which was used to validate the model trained on the bootstrapped sample [16, 17, Raschka, preprint, https://arxiv.org/abs/1811.12808]. Two prediction models were trained: a deep neural network with multiple layers; and a single-layer neural network, the latter of which resembles a simple logistic regression model. We also calculated the Revised Cardiac Risk Index for each patient [18].

Three sequential analyses for MINS and mortality were performed, determined by when variables were available: before surgical referral (age, sex etc); on hospital admission (blood results, intended operation, vital signs etc); subsequent variables (duration of operation, intra-operative blood pressure etc). The association between individual variables and study outcomes was assessed by logistic regression.

We calculated the discrimination of the multiple-layer and single-layer models with the area under the receiver operating characteristic curve, with bootstrap-corrected confidence intervals [16, 19]. The accuracy, specificity and sensitivity of the multiple-layer network were evaluated by dichotomising probabilities (0–1) as no outcome (< 0.5) or outcome (≥ 0.5). We also calculated the area under the receiver operating characteristic curve for subgroups based on age, sex and history of cardiovascular disease (defined as a history of coronary artery disease, heart failure, stroke, atrial fibrillation or cardiac arrest). Calibration curves and calibration errors for MINS were assessed using an isotonic regression to transform model outputs with one-half of the data (out-of-bag), which was tested on the other half of the data [Guo et al., preprint, https://arxiv.org/abs/1706.04599v2]. Permutation tests were used to determine the contributions of clusters of variables to the prediction model (online Supporting Information Appendix S1). We used Python 3 for analyses and fitted and evaluated regression models with the Statsmodel package [20].

Results

Table 2 lists selected pre-operative characteristics of the 24,589 participants, 4030 (16%) of whom were diagnosed with MINS and 406 (2%) of whom died within 30 postoperative days.

Discrimination for MINS by single-layer vs. multiple-layer neural networks generated areas (95%CI) under the receiver operating characteristic curve of: 0.70 (0.69–0.71) vs. 0.71 (0.70–0.73) with variables available before surgical referral, $p < 0.001$; 0.73 (0.72–0.74) vs. 0.75 (0.74–0.76) with additional variables available on admission, but before surgery, $p < 0.001$; and 0.76 (0.75–0.77) vs. 0.77 (0.76–0.78) with the addition of subsequent variables, $p < 0.001$ (online Supporting Information Figure S1). The sequential addition of variables increased the area (95%CI) under the receiver operating characteristic curve for the multiple-layer neural network by 0.036 (0.026–0.043) and 0.025 (0.017–0.035), respectively, $p < 0.001$ for both. Discrimination for MINS by the Revised Cardiac Risk Index was poorer, area (95%CI) under the receiver operating characteristic curve 0.63 (0.61–0.64), $p < 0.001$ when compared with any of the multiple-layer neural networks. The multiple-layer neural network trained using all data predicted MINS with an accuracy of 70% in the validation dataset (Table 3). The calibration error for this network was 19% (online Supporting Information Figure S2).

Discrimination for mortality by single-layer vs. multiple-layer neural networks generated areas (95%CI) under the receiver operating characteristic curve of: 0.71 (0.66–0.76) vs. 0.74 (0.71–0.77) with variables available before surgical referral, $p = 0.04$; 0.78 (0.73–0.82) vs. 0.83 (0.79–0.86) with additional variables available on admission but before surgery, $p = 0.01$; and 0.87 (0.83–0.89) vs. 0.87 (0.85–0.90) with the addition of subsequent variables, $p = 0.52$ (online Supporting Information Figure S3). The sequential addition of variables increased the area (95%CI) under the receiver operating characteristic curve for the multiple-layer neural network by 0.084 (0.037–0.113) and 0.047 (0.021–0.082), respectively, $p < 0.001$ and $p = 0.006$. Discrimination for 30-day death by the Revised Cardiac Risk Index was poorer, area (95%CI) under the receiver operating characteristic curve 0.67 (0.64–0.71), $p < 0.001$ when compared with any of the multiple-layer neural networks. The multiple-layer neural network trained using all data predicted mortality with an accuracy of 89% in the validation dataset (Table 3).

Patient characteristics and pre-existing medical conditions contributed the most in terms of discrimination for MINS in the multiple-layer neural network (Fig. 1). Table 3 lists accuracy, specificity and sensitivity of the multiple-layer neural network when probabilities (0–1) were dichotomised as no outcome (< 0.5) or outcome (≥ 0.5). Subgroup discrimination for myocardial injury by the multiple-layer neural network generated areas (95%CI) under the receiver operating characteristic curve of: 0.78 (0.76–0.79) for women vs. 0.76 (0.75–0.77) for men, $p = 0.06$; 0.72 (0.70–0.75) for history of cardiovascular disease vs. 0.77 (0.76–0.78) for no history of cardiovascular disease, $p < 0.01$; and 0.77 (0.75–0.79) for age 45–59 y vs. 0.75 (0.73–0.76) for age 60–74 y vs. 0.69 (0.67–0.71) for age ≥ 75 years, with $p = 0.03$ and $p < 0.01$ for age 60–74 y and ≥ 75 y vs. age 45–59 y, respectively (online Supporting Information Figure S4).

Discussion

This analysis from a large, representative, international cohort demonstrates that neural networks can be trained to

Table 2 Odds ratio (OR) for associations of selected pre-operative categorical and continuous variables with myocardial injury and death before 30 postoperative days, using logistic and linear regression, respectively. Values are mean (SD) or number (proportion).

			Myocardial injury			Death		
	Missing	Total	No	Yes	OR (95%CI)	No	Yes	OR (95%CI)
Participants		24,589	20,559	4030		24,183	406	
Sex; female	0	11,879 (48%)	10,253 (50%)	1626 (40%)	1.5 (1.4–1.6)	11,709 (48%)	170 (42%)	0.8 (0.6–0.9)
Age; y	1	63.2 (10.8)	62.2 (10.4)	68.4 (11.6)	1.05 (1.05–1.06)	63.2 (10.8)	67.3 (12.0)	1.03 (1.03–1.04)
BMI; kg.m^2	1392	27.5 (7.2)	27.7 (7.4)	26.7 (6.0)	0.97 (0.97–0.98)	27.6 (7.2)	25.0 (6.1)	0.92 (0.90–0.94)
Atrial fibrillation	167	704 (3%)	459 (2%)	235 (6%)	2.7 (2.3–3.2)	682 (3%)	22 (6%)	2.2 (1.4–3.4)
Heart failure	132	686 (3%)	415 (2%)	271 (7%)	3.5 (3.0–4.1)	651 (3%)	35 (10%)	3.8 (2.7–5.4)
Coronary artery disease	127	3309 (14%)	2315 (11%)	994 (25%)	2.6 (2.4–2.8)	3223 (13%)	86 (24%)	2.0 (1.6–2.5)
High-risk coronary artery disease	0	201 (1%)	119 (1%)	82 (2%)	3.6 (2.7–4.7)	189 (1%)	12 (3%)	3.9 (2.1–7.0)
COPD	0	1901 (8%)	1385 (7%)	516 (13%)	2.0 (1.8–2.3)	1837 (8%)	64 (16%)	2.3 (1.7–3.0)
DVT or PE	137	817 (3%)	636 (3%)	181 (5%)	1.5 (1.3–1.8)	799 (3%)	18 (5%)	1.5 (0.9–2.5)
Stroke	0	1451 (6%)	1053 (5%)	388 (10%)	2.0 (1.7–2.2)	1407 (6%)	44 (11%)	2.0 (1.4–2.7)
Diabetes	98	5307 (22%)	4039 (20%)	1268 (32%)	1.9 (1.8–2.0)	5177 (22%)	130 (35%)	1.9 (1.6–2.4)
Hypertension	83	12,358 (50%)	9729 (48%)	2629 (66%)	2.1 (2.0–2.3)	12,125 (50%)	233 (60%)	1.5 (1.2–1.8)
Creatinine; mg.dl^{-1}	1576	1.0 (1.1)	1.0 (0.9)	1.4 (1.6)	1.4 (1.3–1.4)	1.0 (1.0)	1.5 (2.7)	1.1 (1.1–1.2)
Surgery for cancer	1	5238 (21%)	4396 (21%)	842 (21%)	1.0 (0.9–1.1)	5123 (21%)	115 (28%)	1.5 (1.2–1.8)
Surgery for fracture	13	1332 (5%)	1037 (5%)	295 (7%)	1.5 (1.3–1.7)	1306 (5%)	26 (6%)	1.2 (0.8–1.8)
Endoscopic surgery	47	6105 (25%)	5341 (26%)	764 (19%)	0.7 (0.6–0.7)	6052 (25%)	53 (13%)	0.5 (0.3–0.6)
General anaesthesia	28	18,025 (73%)	15,182 (74%)	2843 (71%)	0.85 (0.79–0.91)	17,697 (73%)	328 (81%)	1.5 (1.2–2.0)

COPD, chronic obstructive pulmonary disease; DVT, deep venous thrombosis; PE, pulmonary embolism.

Table 3 The accuracy, specificity and sensitivity of the multiple-layer neural network for myocardial injury and death after dichotomising probabilities (0–1) as no outcome (< 0.5) or outcome (≥ 0.5). Values are proportion (95%CI).

Groups of variables in model	Myocardial injury			Death		
	Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity
First	0.65 (0.62–0.68)	0.65 (0.60–0.69)	0.66 (0.62–0.71)	0.74 (0.66–0.80)	0.74 (0.66–0.80)	0.63 (0.53–0.73)
First and second	0.68 (0.64–0.71)	0.68 (0.63–0.73)	0.69 (0.63–0.75)	0.82 (0.77–0.87)	0.83 (0.77–0.87)	0.67 (0.56–0.78)
First, second and third	0.70 (0.67–0.74)	0.70 (0.65–0.76)	0.70 (0.63–0.76)	0.89 (0.84–0.92)	0.89 (0.84–0.92)	0.69 (0.58–0.79)

predict MINS and death after non-cardiac surgery. Discrimination for these outcomes provided by a single-layer neural network, which resembles a simple logistic regression model, was in most situations similar to that provided by a multiple-layer neural network. Most discriminatory information was provided by pre-operative patient characteristics, with discrimination increased by pre-operative investigations and peri-operative variables.

Our models predicted death more accurately than MINS, a complex outcome that is a sensitive but non-specific marker of peri-operative myocardial stress. Ordinal or continuous measures of peri-operative heart rate, blood pressure and oxygenation, rather than their dichotomisation, might improve the discrimination for MINS. Peri-operative hypotension is common and associated with an increased risk of both MINS and death: interventions designed to limit supply–demand mismatch, which may account for 75% of myocardial ischaemia, offer the greatest opportunity to reduce it, as most pre-operative variables are fixed [3, 21–23]. Machine learning algorithms could continuously update probabilities of myocardial injury and death throughout the peri-operative period, which in turn could be linked with alarms and embedded decision support to guide interventions that might reduce predicted complications.

Machine learning has been used to calculate the probability of postoperative complications, such as acute kidney injury and a range of other complications [24–27]. Discrimination by machine learning models is not always better than discrimination by standard statistical modelling,

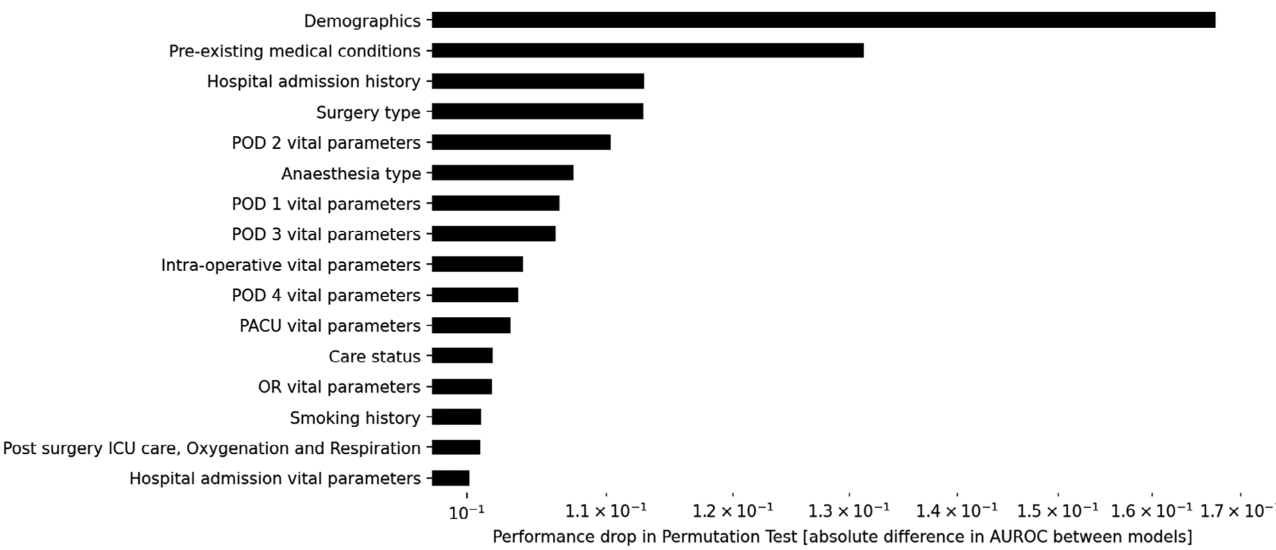

Figure 1 The contributions of clusters of variables to discrimination of postoperative myocardial injury by the multiple-layer neural network (see Table 1). POD, postoperative day; OR, operating room; ICU, intensive care unit; PACU, post-anaesthesia care unit; AUROC, area under the receiving operator characteristic curve

but we think that it is easier to automate these models to be continuously updated by local data than it is standard models [28]. It is unsurprising that the discrimination of the Revised Cardiac Risk Index was less than our models as it was not developed or validated to predict death or MINS [18]. We could not compare the neural networks with other more contemporary risk scores as key components of these were not available. For example, we did not collect ASA physical status, a component of the National Surgical Quality Improvement Program Myocardial Infarction and Cardiac Arrest (NSQIP MICA) calculation and we did not have the administrative data used by the amended Risk Stratification Index [29, 30].

Model calibration depends on the outcome more than discrimination. For example, if a system was designed to identify all patients with a specific risk (e.g. 10%) for an adverse event (i.e. MINS or death), careful calibration would be required. On the other hand, if the system was designed to identify a certain proportion of patients at the highest risk (which may be useful in healthcare settings with constrained resources), calibration is less important since the ranking of risk is the relevant metric. Depending on these requirements, calibration may be prioritised as part of the model training process or performed as a second step with an additional model (in our case by isotonic regression). The latter does not exclude the live output of calibrated predictions. The only necessity for such a real-time approach is the systematic, automated and instantaneous collection of patient data.

It is impossible to validate our neural networks externally in the absence of another large dataset with routinely collected troponin assays and clinical information necessary to diagnose MINS. The aim of our study was to assess the ability of machine learning, combined with standardised data collection, to predict MINS and early postoperative death. The intention was not to construct a fixed model for general use but an algorithm that can be continuously and prospectively tested, adapted and improved using local data. Although we did not compare the performance of our neural network models directly with a standard statistical model, the single-layer network has characteristics that are similar to a conventional model. Likewise, we did not assess other machine learning techniques, which might perform similarly to neural networks [31].

In summary, we have reported discrimination, calibration, accuracy, sensitivity and specificity for MINS and death within 30 days of non-cardiac surgery using single and multiple-layer neural networks.

Acknowledgements

The VISION study was registered (NCT00512109). JN and MS are joint first authors and PD and GH are joint senior authors. MS is supported by the Royal Perth Hospital Medical Research Foundation and has received fees or support from Medtronic, Abbott, ReCor, Novartis, Servier, Pfizer and Boehringer-Ingelheim. DS is a consultant for Edwards LifeSciences (Irvine, CA, USA) and serves on advisory boards and has equity interests in Health Data Analytics Institute (Boston, MA, USA, Sensifree (Cupertino, CA, USA) and Perceptive Medical (Newport Beach, CA, USA). AM is the recipient of an Australian Research Council Future Fellowship Award. PD has provided consulting services for Roche Diagnostics and Trimedics and has served on advisory boards for Bayer and Quidel Canada. He has received research support from Abbott Diagnostics, Cloud DX, Philips Healthcare, Roche Diagnostics and Siemens. CC is supported by an Australian National Health and Medical Research Council Investigator Award. FB is a recipient of a Research Early Career Award from Hamilton Health Sciences and has received grants from Roche and Siemens, unrelated to this work. PM is funded by an Australian National Health and Medical Research Council Leadership Fellowship. NM is supported by a Chair Award, Programme Grant and Research Excellence Award from the British Heart Foundation. GH is a recipient of a Western Australian Health Research Excellence Award. Open access publishing facilitated by The University of Western Australia, as part of the Wiley - The University of Western Australia agreement via the Council of Australian University Librarians.

References

1. Meara JG, Leather AJ, Hagander L, et al. Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. *Lancet* 2015; **386**: 569–624.
2. Devereaux PJ, Sessler DI. Cardiac complications in patients undergoing major noncardiac surgery. *New England Journal of Medicine* 2015; **373**: 2258–69.
3. Devereaux PJ, Szczeklik W. Myocardial injury after non-cardiac surgery: diagnosis and management. *European Heart Journal* 2020; **41**: 3083–91.
4. Devereaux PJ, Biccard BM, Sigamani A, et al. Association of postoperative high-sensitivity troponin levels with myocardial injury and 30-day mortality among patients undergoing noncardiac surgery. *Journal of the American Medical Association* 2017; **317**: 1642–51.
5. Devereaux PJ, Chan MT, Alonso-Coello P, et al. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. *Journal of the American Medical Association* 2012; **307**: 2295–304.
6. Devereaux PJ, Duceppe E, Guyatt G, et al. Dabigatran in patients with myocardial injury after non-cardiac surgery (MANAGE): an international, randomised, placebo-controlled trial. *Lancet* 2018; **391**: 2325–34.
7. Beattie WS, Wijeysundera DN, Chan MTV, et al. Survival after isolated post-operative troponin elevation. *Journal of the American College of Cardiology* 2017; **70**: 907–8.
8. Beattie WS, Wijeysundera DN, Chan MTV, et al. Implication of major adverse postoperative events and myocardial injury on disability and survival: a planned subanalysis of the ENIGMA-II trial. *Anesthesia and Analgesia* 2018; **127**: 1118–26.
9. Botto F, Alonso-Coello P, Chan MT. Myocardial injury after noncardiac surgery: a large, international, prospective cohort study establishing diagnostic criteria, characteristics, predictors, and 30-day outcomes. *Anesthesiology* 2014; **120**: 564–78.
10. Padmanabhan S, Tran TQB, Dominiczak AF. Artificial Intelligence in hypertension: seeing through a glass darkly. *Circulation Research* 2021; **128**: 1100–18.
11. Pedregosa F, Varoquaux GG, Gramfort A, Michel V, et al. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research* 2011; **12**: 2825–30.
12. TensorFlow. https://www.tensorflow.org/about/bib (accessed 23/02/2023).
13. Usenix. TensorFlow: A System for Large-Scale Machine Learning. 2016. https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf (accessed 23/02/2023).
14. Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation* 1998; **10**: 1895–923.
15. Vanwinckelen G, Blockeel H. On estimating model accuracy with repeated cross-validation. *BeneLearn 2012: Proceedings of the 21st Belgian-Dutch Conference on Machine Learning*. 2012: 39–44.
16. Efron B, Tibshirani RJ. Better bootstrap confidence intervals. In: Efron B, Tibshirani RJ, eds. *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993: 178–99.
17. Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association* 1983; **78**: 316–31.
18. Lee TH, Marcantonio ER, Mangione CM, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation* 1999; **100**: 1043–9.
19. Jung K, Lee J, Gupta V, Cho G. Comparison of bootstrap confidence interval methods for GSCA using a Monte Carlo simulation. *Frontiers in Psychology* 2019; **10**: 2215.
20. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. *The 9th Python in Science Conference*. 2010. https://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf (accessed 22/03/2023).
21. Abbott TEF, Pearse RM, Archbold RA, et al. A prospective international multicentre cohort study of intraoperative heart rate and systolic blood pressure and myocardial injury after noncardiac surgery: results of the VISION study. *Anesthesia and Analgesia* 2018; **126**: 1936–45.
22. van Lier F, Wesdorp F, Liem VGB, et al. Association between postoperative mean arterial blood pressure and myocardial injury after noncardiac surgery. *British Journal of Anaesthesia* 2018; **120**: 77–83.
23. Sessler DI, Meyhoff CS, Zimmerman NM, et al. Period-dependent associations between hypotension during and for four days after noncardiac surgery and a composite of myocardial infarction and death: a substudy of the POISE-2 trial. *Anesthesiology* 2018; **128**: 317–27.
24. Filiberto AC, Ozrazgat-Baslanti T, Loftus TJ, et al. Optimizing predictive strategies for acute kidney injury after major vascular surgery. *Surgery* 2021; **170**: 298–303.
25. Adhikari L, Ozrazgat-Baslanti T, Ruppert M, et al. Improved predictive models for acute kidney injury with IDEA: intraoperative data embedded analytics. *PLoS One* 2019; **14**: e0214904.
26. Xue B, Li D, Lu C, et al. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. *Journal of the American Medical Association Network Open* 2021; **4**: e212240.
27. Bihorac A, Ozrazgat-Baslanti T, Ebadi A, et al. MySurgeryRisk: development and validation of a machine-learning risk algorithm for major complications and death after surgery. *Annals of Surgery* 2019; **269**: 652–62.
28. Racine AM, Tommet D, D'Aquila ML, et al. Machine learning to develop and internally validate a predictive model for post-operative delirium in a prospective, observational clinical

cohort study of older surgical patients. *Journal of General Internal Medicine* 2021; **36**: 265–73.
29. Gupta PK, Gupta H, Sundaram A, et al. Development and validation of a risk calculator for prediction of cardiac risk after surgery. *Circulation* 2011; **124**: 381–7.
30. Chamoun GF, Li L, Chamoun NG, Saini V, Sessler DI. Comparison of an updated risk stratification index to hierarchical condition categories. *Anesthesiology* 2018; **128**: 109–16.
31. Greenwald S, Chamoun GF, Chamoun NG, et al. Risk stratification index 3.0, a broad set of models for predicting adverse events during and after hospital admission. *Anesthesiology* 2022; **137**: 673–86.

Supporting Information

Additional supporting information may be found online via the journal website.

Appendix S1. Design and training of neural networks and permutation testing.

Figure S1. Multiple-layer neural network-based prediction of MINS following non-cardiac surgery using data from different phases of assessment.

Figure S2. Calibration plots for the probabilities of MINS.

Figure S3. Multiple-layer neural network-based prediction of 30-day death following non-cardiac surgery using data from different phases of assessment.

Figure S4. Receiver operating characteristic curves for subgroups.

