Electrocardiographic deep learning for predicting post-procedural mortality: a model development and validation study

David Ouyang, John Theurer, Nathan R Stein, J Weston Hughes, Pierre Elias, Bryan He, Neal Yuan, Grant Duffy, Roopinder K Sandhu, Joseph Ebinger, Patrick Botting, Melvin Jujjavarapu, Brian Claggett, James E Tooley, Tim Poterucha, Jonathan H Chen, Michael Nurok, Marco Perez, Adler Perotte, James Y Zou, Nancy R Cook, Sumeet S Chugh, Susan Cheng, Christine M Albert

Summary

Background Preoperative risk assessments used in clinical practice are insufficient in their ability to identify risk for postoperative mortality. Deep-learning analysis of electrocardiography can identify hidden risk markers that can help to prognosticate postoperative mortality. We aimed to develop a prognostic model that accurately predicts postoperative mortality in patients undergoing medical procedures and who had received preoperative electrocardiographic diagnostic testing.

Methods In a derivation cohort of preoperative patients with available electrocardiograms (ECGs) from Cedars-Sinai Medical Center (Los Angeles, CA, USA) between Jan 1, 2015 and Dec 31, 2019, a deep-learning algorithm was developed to leverage waveform signals to discriminate postoperative mortality. We randomly split patients (8:1:1) into subsets for training, internal validation, and final algorithm test analyses. Model performance was assessed using area under the receiver operating characteristic curve (AUC) values in the hold-out test dataset and in two external hospital cohorts and compared with the established Revised Cardiac Risk Index (RCRI) score. The primary outcome was post-procedural mortality across three health-care systems.

Findings 45 969 patients had a complete ECG waveform image available for at least one 12-lead ECG performed within the 30 days before the procedure date (59 975 inpatient procedures and 112 794 ECGs): 36 839 patients in the training dataset, 4549 in the internal validation dataset, and 4581 in the internal test dataset. In the held-out internal test cohort, the algorithm discriminates mortality with an AUC value of 0·83 (95% CI 0·79–0·87), surpassing the discrimination of the RCRI score with an AUC of 0·67 (0·61–0·72). The algorithm similarly discriminated risk for mortality in two independent US health-care systems, with AUCs of 0·79 (0·75–0·83) and 0·75 (0·74–0·76), respectively. Patients determined to be high risk by the deep-learning model had an unadjusted odds ratio (OR) of 8·83 (5·57–13·20) for postoperative mortality compared with an unadjusted OR of 2·08 (0·77–3·50) for postoperative mortality for RCRI scores of more than 2. The deep-learning algorithm performed similarly for patients undergoing cardiac surgery (AUC 0·85 [0·77–0·92]), non-cardiac surgery (AUC 0·83 [0·79–0·88]), and catheterisation or endoscopy suite procedures (AUC 0·76 [0·72–0·81]).

Interpretation A deep-learning algorithm interpreting preoperative ECGs can improve discrimination of postoperative mortality. The deep-learning algorithm worked equally well for risk stratification of cardiac surgeries, non-cardiac surgeries, and catheterisation laboratory procedures, and was validated in three independent health-care systems. This algorithm can provide additional information to clinicians making the decision to perform medical procedures and stratify the risk of future complications.

Funding National Heart, Lung, and Blood Institute.

Introduction

In the USA, more than 20 million surgeries are performed annually.[1] Preoperative risk assessment for adverse procedural outcomes—the most serious of which include death—is routinely performed in clinical practice.[2,3] However, current approaches to predicting postoperative risk remain inadequate.[4–6] Over the past three decades, expert guidelines and tools for facilitating preoperative assessments have evolved to include biomarkers and demographic and clinical data.[6–8] However, even the most comprehensive risk scores based on recognised risk markers provide only a modest ability to discriminate postoperative outcomes, with areas under the receiver operating characteristic curve (AUC) ranging from 0·57 to 0·75.[6,9–12]

Early identification of patients at high risk for post-procedural mortality can guide patient care and the consideration of alternative treatment pathways, and aid in

Lancet Digit Health 2024; 6: e70–78

Published Online December 7, 2023 https://doi.org/10.1016/S2589-7500(23)00220-0

See Comment page e4

Department of Cardiology, Smidt Heart Institute (D Ouyang MD, J Theurer BS, N R Stein MD, N Yuan MD, G Duffy BS, R K Sandhu MD, J Ebinger MD, P Botting BS, M Jujjavarapu BS, S S Chugh MD, S Cheng MD, C M Albert MD), **Division of Artificial Intelligence in Medicine, Department of Medicine** (D Ouyang, S S Chugh), **and Division of Anesthesia, Department of Surgery** (M Nurok MD), **Cedars-Sinai Medical Center, Los Angeles, CA, USA; Department of Computer Science** (J W Hughes BS, B He PhD, J Y Zou PhD), **Division of Cardiology** (J E Tooley MD, M Perez MD) **and Division of Bioinformatics Research** (J H Chen MD PhD), **Department of Medicine, and Department of Biomedical Data Science** (J Y Zou), **Stanford University, Palo Alto, CA, USA; Milstein Division of Cardiology, Department of Medicine** (P Elias MD, T Poterucha MD, A Perotte PhD) **and Department of Biomedical Informatics** (P Elias), **Columbia University Irving Medical Center, New York, NY, USA; Division of Cardiovascular Medicine** (B Claggett PhD) **and Division of Preventive Medicine** (N R Cook PhD), **Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA**

Correspondence to:
Dr David Ouyang, Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90035, USA
david.ouyang@cshs.org

Research in context

Evidence before this study

Early identification of patients at high risk for post-procedural mortality can guide patient care and the consideration of alternative treatment pathways, and aid in shared decision making and communication of risk to patients. We did a search of PubMed for society guidelines and primary research articles to inform our understanding of the field. We did not do a formal meta-analysis but relevant literature from English-language publications was reviewed. We found that postoperative biomarker levels have been used to risk-stratify patients for postoperative mortality, with intermediate discrimination ability. However, even the most comprehensive risk scores based on recognised risk markers provide only a modest ability to discriminate postoperative outcomes. All available literature was reviewed between July 1, 2021 and Jan 1, 2023 before publication of this Article.

Added value of this study

Deep-learning approaches applied to preoperative electrocardiograms (ECGs) can offer additional information that could improve discrimination of postoperative mortality outcomes and improve on current preoperative risk-stratification tools. PreOpNet is a deep-learning algorithm that leverages ECG waveform signals to augment discrimination of postoperative mortality outcomes, and its performance was evaluated in three independent health-care systems.

Implications of all the available evidence

ECGs contain a lot of information that can assist in preoperative risk stratification, and the evaluation of ECGs using deep learning can discriminate postoperative mortality better than when using the established Revised Cardiac Risk Index.

shared decision making and communication of risk to patients.[13,14] Post-operative biomarker levels have been used to stratify patients according to risk of postoperative mortality, with strong discrimination;[15–17] however, the dependence on post-procedural information limits the opportunity to risk-stratify preoperatively. Novel methods of perioperative risk assessment are needed to achieve better discrimination across the heterogeneous population of preoperative patients than is currently possible.

The emergence of deep-learning analyses now offers the opportunity to capture previously unmeasured risk markers and simultaneously assess complex, interactive relationships from readily available clinical resources for risk prediction.[18–20] One such ideal resource for perioperative risk discrimination is the 12-lead electrocardiogram (ECG). ECGs are inexpensive, non-invasive, and rapid diagnostic tests that are routinely obtained in the preoperative setting as per clinical guidelines.[21,22] Previous studies have applied deep-learning algorithms to ECG waveforms to identify clinical traits and outcomes not previously associated with conventional ECG measures or even expert human ECG interpretations.[3,23–27] Therefore, we hypothesised that deep-learning algorithms applied to a single preoperative ECG could reliably discriminate postoperative mortality outcomes and improve upon established clinical approaches to preoperative assessment. To evaluate this hypothesis, we did a comprehensive study of an artificial intelligence (AI) algorithm trained on perioperative ECGs and evaluated the performance of the resulting model on cohorts from three independent health-care systems.

Methods

Derivation cohort

For this model development study, all patients undergoing inpatient procedures at Cedars-Sinai Medical Center (CSMC) between Jan 1, 2015 and Dec 31, 2019 were included in the derivation cohort. Of this source population, we included in the analysis all patients who had a complete ECG waveform image available for at least one 12-lead ECG performed within the 30 days before the procedure date. From this derivation cohort, we randomly split patients (8:1:1) into subsets for training, internal validation, and final algorithm test analyses using a computer-generated random sequence. All ECG waveform data were acquired through the clinical enterprise data warehouse at a sampling rate of 500 Hz and extracted as 12×5000 matrices of amplitude values corresponding to a 10 s period. We excluded ECGs with missing leads from the analyses. We also obtained associated clinical data for each patient from electronic health records. The study was approved by the Institutional Review Boards of Cedars-Sinai Medical Center, Stanford University, and Columbia University. Exemption from informed consent was obtained given the use of retrospective, de-identified data for the study. We used the TRIPOD reporting guidelines for this Article.

See **Online** for appendix

Clinical and outcome assessments

Patients' demographic, clinical, and outcome data were assessed from electronic health records at the time of each procedure. From these data, the preoperative clinical characteristics needed for calculating the Revised Cardiac Risk Index (RCRI)[13] were identified, including coronary artery disease, congestive heart failure, stroke or transient ischaemia attack, preoperative insulin use, creatinine concentrations higher than 2 mg/dL, and surgery with increased risk of complications as defined by American College of Cardiology and American Heart Association guidelines.[2] For the main analysis, the outcome was death during hospitalisation or during readmission within 30 days. Procedures were classified by type, including procedures performed in the operating room by cardiac surgeons (cardiac surgeries), procedures performed in the operating room by other surgeons (non-cardiac surgeries; appendix p 1), and procedures performed in the catheterisation laboratory or endoscopy suite. Procedural complications were identified using relevant postoperative diagnoses that were present for the first time after the procedure date or at discharge. Outcomes were adjudicated up to 30 days after the date of procedure, with multiple procedures having independent outcome windows on the basis of procedure date. If the same outcome fell within the 30-day window for multiple procedures, the outcome was attributed to each of the procedures. Diagnoses were encoded by International Classification of Diseases (ICD)-9 or ICD-10 codes, and major adverse cardiovascular events (MACE) were identified using previously validated criteria from the electronic health record.[28,29]

ECG assessments

We trained and validated a deep-learning algorithm based on waveform signals from a single preoperative 12-lead ECG, termed PreOpNet, on the outcome of postoperative mortality (figure 1). The input of the model was a 12-lead ECG obtained within the 30 days before an operative procedure, and the outputs were the hospitalisation-level outcomes following that procedure. Patients with multiple procedures were treated independently during model training, with each ECG paired with the most proximal subsequent procedure. To maximise sample size, similar to previous studies of deep learning using electrocardiography, all ECGs in the window of interest were used in the training set as training examples. In the internal validation and test cohorts, the analysis was limited to a single ECG (ie, the most proximal to the procedure of interest for an individual patient) to mimic how the model would be applied in clinical practice. When used, clinical features were input into the last fully connected layer before model output. Models were trained using the PyTorch deep-learning library.

Deep-learning algorithm development

Based on previous literature regarding lightweight, deep-learning model architecture design and neural architecture search, PreOpNet was designed to analyse 12-lead ECG waveform data starting with atrous convolutions followed by subsequent multi-channel one-dimensional convolutions. The number of layers paralleled the design of EfficientNet and were limited to less than a tenth of the size of previously described architectures to optimise model run-time and minimise model complexity.[3–5] After initial atrous layers, PreOpNet incorporated convolutional layers with an inverted residual structure, for which the input and output are bottleneck layers with an intermediate expansion layer. In each set of expansion layers with bottleneck layers preceding and succeeding, the number of input channels gradually increased to allow for the integration of information across ECG leads. The deep-learning model had only the 12-lead ECG waveform data as inputs, and

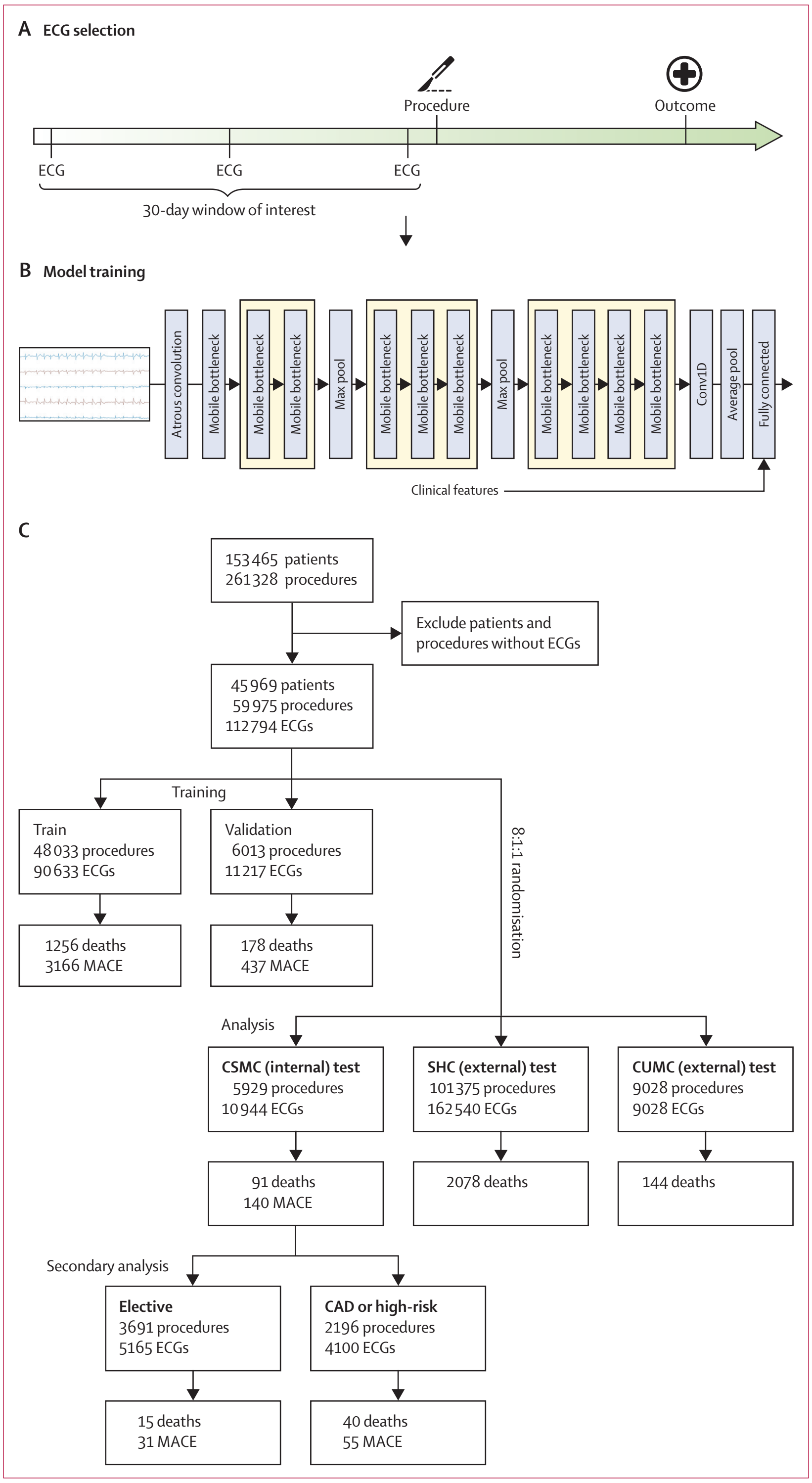

Figure 1: **CSMC cohort sampling**
(A) ECGs within 30 days of an inpatient procedure were selected for the study and paired with postoperative outcomes. (B) A novel, lightweight model architecture was trained to discriminate postoperative mortality outcomes and complications, with input of the nearest 12-lead ECG. (C) Participant selection flow chart. CAD=coronary artery disease. CSMC=Cedars-Sinai Medical Center. CUMC=Columbia University Medical Center. ECG=electrocardiogram. MACE=major adverse cardiovascular events. SHC=Stanford Healthcare.

comparison, non-deep-learning models included clinical data, such as age and sex, as well as structured ECG information.

We initialised the model with random weights and trained with a loss function of binary cross-entropy for 100 epochs using an Adam optimiser with an initial learning rate between 0·005 and 0·0001. Early stopping was done on the basis of the validation dataset's AUC. In the process of hyperparameter tuning, we grid-searched the atrous convolution's dilation and step-size for optimal

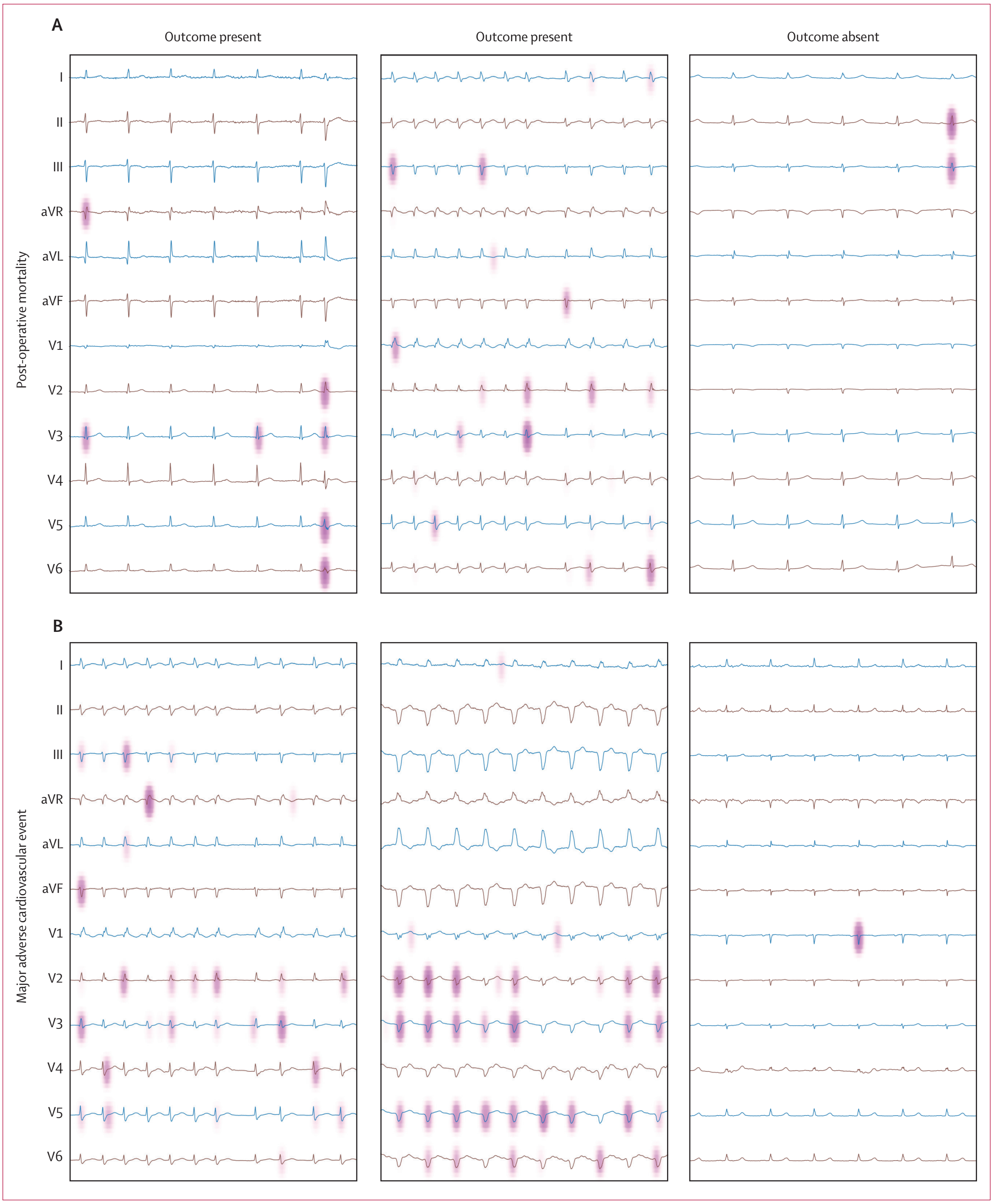

***Figure 2:* Interpretability analysis of PreOpNet**

Select electrocardiograms before procedures with positive and negative outcomes are presented, highlighting the most relevant features as determined by interpretability analysis: discrimination of mortality and major adverse cardiovascular events.

AUC with all other hyperparameters held constant (appendix p 6). We note that our new model architecture is less than a 50th of the size of previous models and efficient enough to be run on a standard computer (appendix p 2). Local, interpretable, model-agnostic explanations were used to identify and visualise relevant features in the ECG used for model decision making (figure 2).

Evaluation of model performance and comparison with an established risk calculator

Following algorithm development in the training and internal validation datasets, the ability of PreOpNet to discriminate the primary outcome of postoperative mortality in the held-out internal test dataset was assessed using AUC values. To understand the model's performance in key patient populations, we did independent analyses that were limited to either: (1) patients undergoing cardiac surgery, (2) non-cardiac surgery, or (3) interventional endoscopy suite or catheterisation laboratory procedures; or (4) patients with known cardiovascular disease or undergoing intermediate-risk or high-risk surgeries. We compared the performance of PreOpNet in the held-out internal test dataset with an established risk calculator (using the RCRI score),[6] conventional ECG measures and interpretations, and alternative algorithms based on ECG measurement data.

The association of PreOpNet with perioperative outcomes was compared with that of an RCRI score >2, a commonly used threshold to identify high perioperative risk associated with a MACE rate greater than 8% in meta-analyses. In the training set, 15% of the cohort had an RCRI score >2; thus, we set a similar threshold for high perioperative risk (ie, the top 15%) for PreOpNet. We estimated odds ratios (ORs), sensitivity, and specificity for postoperative mortality and MACE at this threshold. In addition, we did secondary analyses with MACE as the outcome. We calculated the continuous and categorical net reclassification index (NRI) for mortality and MACE rate associated with the addition of PreOpNet to an RCRI score >2, and used 10 000 bootstrapped samples to obtain 95% CIs for each estimate. To discern the potentially most informative features of the ECG waveform, in the context of comparing performance to that of the RCRI, 0·5% of the waveform for 1000 samples per study were iteratively randomly perturbed to identify which changes most affected model performance. We evaluated implementation timings using Python's timeit module.

External validation of PreOpNet for discrimination of postoperative mortality

To assess algorithm performance in other hospital settings, the PreOpNet algorithm was applied without any additional further fine-tuning or training to patients from two separate external health-care systems. We did not transfer patient data across institutions to maximise the rigour of external validation; instead, investigators from each external institution independently collected information on outcomes and ran model inference on their datasets to report summary statistics. The Stanford Healthcare (SHC) cohort included patients from May 1, 2007 to June 30, 2018. All data on clinical characteristics and outcomes were provided through the Stanford Research Data Repository Observational Medical Outcomes Partnership common data model. ECG waveform data were provided through the TraceMaster data management system (Philips Healthcare; Boston, MA, USA) and pre-processed with a low-pass filter to correct for wandering baselines and normalisation of waveform data. The Columbia University Medical Center (CUMC) cohort included patients from Jan 1 to March 31, 2020. At CUMC, data on clinical characteristics and outcomes were obtained from the clinical enterprise data warehouse and ECG waveform data were obtained from the Muse data management system (GE Healthcare, Chicago, IL, USA). For each of the two external cohorts, the AUC for postoperative mortality from analyses of a single preoperative ECG was calculated. Procedures were linked to the most proximal preceding ECG performed within 30 days before, and postoperative mortality was assessed as mortality within 30 days of the procedure or during hospitalisation (appendix p 1).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Between Jan 1, 2015 and Dec 31, 2019, 153 465 patients aged 18 years or older in the CSMC underwent 261 328 procedures in the operating room, catheterisation laboratory, or endoscopy suite. Of these individuals, 45 969 patients had a complete ECG waveform image available for at least one 12-lead ECG performed within the 30 days before the procedure date (corresponding to 59 975 inpatient procedures and contributing 112 794 ECGs; figure 1). After randomisation, there were 36 839 patients (contributing 90 633 ECGs) in the training dataset, 4549 patients (contributing 11 217 ECGs) in the internal validation dataset, and the remaining 4581 patients (contributing 10 944 ECGs) in the internal test dataset (table 1).

The mean age at the time of preoperative ECG was 65·1 years (SD 15·9), 45·1% were women, and 21·7% had pre-existing coronary artery disease (table 1). There were 1065 (1·8%) subsequent deaths and 1730 (2·9%) post-procedural MACE events during hospitalisation in the cohort (table 1). Compared with patients who were excluded on the basis of not having ECG data within 30 days of their procedure, patients with preoperative ECGs were more likely to be older, male, and with more cardiovascular risk factors (appendix p 3).

For the outcome of mortality, the PreOpNet algorithm was developed in the derivation cohort using training,

	Training subcohort	Validation subcohort	Test subcohort
Patients, n	36 839	4549	4581
Procedures, n	48 033	6013	5929
Electrocardiograms, n	90 633	11 217	10 944
Demographic and clinical characteristics			
Age, years (SD)	65·2 (15·8)	65·0 (16·4)	64·6 (16·1)
Female	21 744 (45·3%)	2706 (45·0%)	2641 (44·6%)
Male	26 289 (54·7%)	3307 (55·0%)	3288 (55·4%)
Caucasian	34 412 (71·6%)	4255 (70·8%)	4143 (69·9%)
Black	5323 (11·1%)	662 (11·0%)	739 (12·5%)
Asian	3298 (6·9%)	406 (6·7%)	401 (6·8%)
Other	4998 (9·9%)	652 (10·5%)	685 (11·6%)
Heart failure	8689 (18·1%)	1192 (19·8%)	1119 (18·9%)
Diabetes	7880 (16·4%)	998 (16·6%)	1061 (17·9%)
Hypertension	16 638 (34·6%)	2187 (36·4%)	2063 (34·8%)
Coronary artery disease	10 421 (21·7%)	1293 (21·5%)	1305 (22·0%)
Stroke	2697 (5·6%)	352 (5·9%)	343 (5·8%)
Renal disease	4475 (9·3%)	640 (10·6%)	568 (9·6%)
Procedure types			
Cardiovascular	19 840 (41·3%)	2595 (43·2%)	2447 (41·3%)
Intraperitoneal, intrathoracic, or suprainguinal vascular	8679 (18·1%)	1074 (17·9%)	1090 (18·4%)
Insulin use before admission	3022 (6·3%)	366 (6·1%)	374 (6·3%)
Creatinine >2 mg/dL	4408 (9·2%)	637 (10·6%)	557 (9·4%)
RCRI >2	2784 (5·8%)	327 (5·4%)	378 (6·4%)
Postoperative outcomes			
Death during hospitalisation	865 (1·8%)	109 (1·8%)	91 (1·5%)
Cardiovascular death	21 (<0·1%)	7 (0·1%)	4 (0·1%)
Major adverse cardiovascular events	1400 (2·9%)	190 (3·2%)	140 (2·4%)
Cardiac arrest	182 (0·4%)	27 (0·4%)	18 (0·3%)
Myocardial infarction	203 (0·4%)	30 (0·5%)	20 (0·3%)
Heart block	129 (0·3%)	22 (0·4%)	9 (0·2%)
Pulmonary oedema	40 (0·1%)	4 (0·1%)	4 (0·1%)

Data are n (%), unless otherwise stated. RCRI=Revised Cardiac Risk Index.

***Table 1:* Characteristics and outcomes of the derivation cohort**

validation, and test datasets (figure 3). In the held-out test dataset, the algorithm was then shown to discriminate mortality with an AUC of 0·83 (95% CI 0·79–0·87; table 2). By contrast, the conventional RCRI score discriminated postoperative mortality with an AUC of 0·67 (0·61–0·72). In the 15% of patients who had an RCRI score of 2 or higher, the unadjusted OR for postoperative mortality was 2·08 (95% CI 0·77–3·50). In comparison, patients in the top 15% of the PreOpNet algorithm had an adjusted OR of 9·17 (95% CI 5·85–13·82) for postoperative mortality (appendix p 5). The addition of the components of the RCRI score to the PreOpNet algorithm trained only on 12-lead ECG waveforms did not significantly improve model performance (AUC 0·83 [95% CI 0·77–0·89]) in the test dataset (appendix p 7). There were no significant differences in model performance in patient subsets defined by age, sex, or race. At the pre-specified calibrated threshold of risk comparable to an RCRI score of 2 or higher, the PreOpNet algorithm demonstrated a specificity of 0·87 (0·86–0·88) and sensitivity of 0·57 (0·48–0·68) for postoperative mortality (appendix p 5). In comparison, the RCRI score of 2 or higher had slightly higher specificity (0·94 [0·93–0·94]) but much lower sensitivity (0·12 [0·05–0·19]). At this threshold, the positive predictive value was slightly higher for PreOpNet than for the RCRI score (table 2).

PreOpNet performed well both in patients undergoing major surgical procedures in the operating room and in patients undergoing procedures in the catheterisation laboratory or endoscopy suite. For patients with surgeries in the operating room, PreOpNet discriminated postoperative mortality with an AUC of 0·84 (0·76–0·92), compared with an AUC of 0·70 (0·61–0·78) for the RCRI score (appendix p 7). For patients with procedures in the catheterisation laboratory or endoscopy suite, PreOpNet discriminated postoperative mortality with an AUC of 0·76 [0·72–0·81]), compared with an AUC of 0·66 (0·60–0·72) for the RCRI score (appendix p 7). PreOpNet performed similarly in discriminating mortality in patients undergoing cardiovascular surgery (AUC 0·85 [0·77–0·92]) and patients undergoing non-cardiac surgery (AUC 0·83 [0·79–0·88]; appendix p 7). The RCRI score discriminated postoperative mortality with an AUC of 0·62 (0·52–0·72) in patients undergoing cardiac surgery and an AUC of 0·70 (0·63–0·77) in patients undergoing non-cardiac surgery (appendix p 7).

Given that ECGs are often not obtained in low-risk patients undergoing low-risk procedures, a secondary analysis was performed in patients most likely to be considered at least moderate-risk (patients either with known cardiovascular disease or those undergoing elective intermediate-risk or high-risk surgery). Without additional subset-specific fine-tuning, the PreOpNet algorithm discriminated postoperative mortality in those considered to be at least moderate-risk with an AUC of 0·80 (0·71–0·88; appendix p 7). In clinical practice, preoperative risk assessment most commonly occurs in the elective procedural setting. Thus, secondary analyses were performed that were limited to those patients in the CSMC test cohort who were undergoing elective procedures (3691 patients contributing 5165 ECGs). In this setting, the PreOpNet algorithm discriminated postoperative mortality with an AUC of 0·80 (0·67–0·92; appendix p 7).

To assess the external validity of the PreOpNet algorithm, its discriminatory ability was evaluated in two external health system cohorts. The external test evaluation cohorts included 101 375 patients in the SHC system, contributing 162 540 ECGs, and 9028 patients in the CUMC system, contributing 9028 ECGs. In the SHC cohort, the postoperative mortality rate was 1·3% and PreOpNet discriminated this outcome with an AUC of 0·75 (95% CI 0·74–0·76). In the CUMC cohort, the postoperative mortality rate was 1·6% and the algorithm discriminated this outcome with an AUC of 0·79 (0·75–0·83; table 2). The PreOpNet algorithm pre-specified high-risk group (>15%) had an unadjusted OR of 5·88 (5·00–7·00) in SHC and 6·20 (3·87–10·41) in CUMC for mortality. Results from analyses of specificity, sensitivity, and positive and negative predictive value were similar in the external validation cohorts when compared with results observed in the CSMC cohort (table 2).

Secondary analyses were performed within the internal test dataset, which used a combined secondary MACE outcome that included non-fatal MACE and postoperative mortality. For this secondary outcome, the PreOpNet algorithm discriminated events in the held-out internal test dataset with an AUC of 0·77 (95% CI 0·73–0·80), whereas the RCRI score had an AUC of 0·63 (0·59–0·68; table 2). Patients with an RCRI score of 2 or higher had an unadjusted OR of 1·67 (0·77–2·68) for MACE when compared with those with an RCRI score less than 2. By contrast, patients identified by the PreOpNet algorithm to be high-risk had an unadjusted OR of 5·38 (3·75–7·49) for MACE (table 2). The PreOpNet algorithm demonstrated a specificity of 0·88 (0·88–0·89) and sensitivity of 0·41 (0·33–0·49), and the RCRI score again had higher specificity (0·94 [0·93–0·94]) but much lower sensitivity (0·10 [0·05–0·15]). The positive predictive value for MACE was higher for PreOpNet than for the RCRI, with similar negative predictive values (table 2).

The ability of the PreOpNet algorithm to reclassify risk in the internal hold-out test dataset was also evaluated. When compared with the RCRI score, application of the PreOpNet algorithm led to significant improvement in the continuous NRI (0·53 [95% CI 0·38 to 0·68]). In categorical analyses using the pre-specified threshold of risk (top 15% PreOpNet prediction), 981 (82·4%) of 1190 patients originally classified as high-risk for MACE by the RCRI score (RCRI ≥2) were identified as low-risk by PreOpNet (appendix p 4). Of these high-to-low-risk reclassified patients, 33 (3·4%) experienced a MACE. By contrast, of the 4739 patients classified as low-risk by RCRI (RCRI <2), the PreOpNet algorithm reclassified 327 (6·9%) as high-risk; of these patients, 26 (8·0%) experienced a MACE. Despite a fair amount of reclassification, the categorical NRI at this cutoff point was not significant for MACE (0·06 [95% CI –0·04 to 0·18]).

To compare the performance of our architecture with other deep-learning models, we implemented a recently described model for ECG waveforms to perform the same task and trained this benchmark model with the same training data with labels of postoperative outcomes. The previously published deep-learning architecture had an AUC of 0·68 (95% CI 0·65–0·70) for predicting death and an AUC of 0·57 (0·55–0·59) for predicting MACE.

In addition to superior performance in predicting mortality and complications, PreOpNet is a highly efficient deep-learning model architecture with fewer parameters and requiring less computational power to train (up to 100 times smaller) than other published architectures (appendix p 2). The improvement in speedup and computational efficiency allows for the model to be run solely on a standard central processing unit at inference time, which allows for a web interface deployment for ready access by clinicians (appendix p 7). In a series of repeated run-time experiments, the PreOpNet application accessed on a standard clinical workstation (Windows 10 64-bit operating system, 3·8 GHz processor) was able to take in image data from 50 de-novo ECGs and output postoperative risk estimates within 0·032 s (SD 0·004) per ECG for the local software installation and 0·041 s (SD 0·006) per ECG for the web application accessed by a mobile phone.

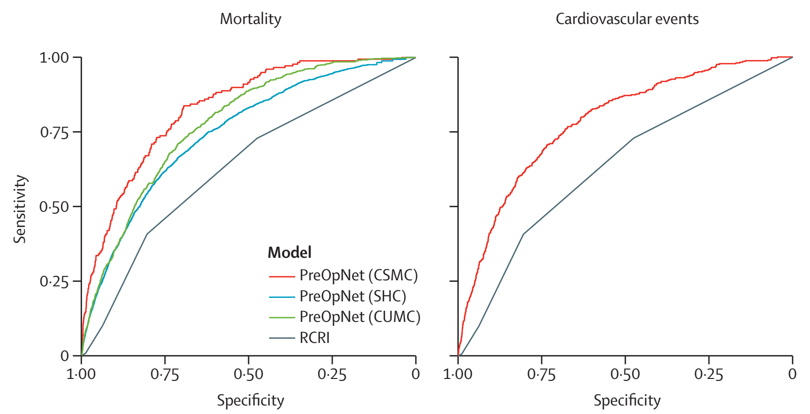

Figure 3: **PreOpNet workflow and results**
Performance of PreOpNet at CSMC, SHC, and CUMC, and RCRI in discriminating postoperative mortality and major adverse cardiovascular events. CSMC=Cedars-Sinai Medical Center. CUMC=Columbia University Medical Center. RCRI=Revised Cardiac Risk Index. SHC=Stanford Healthcare.

	AUC	Specificity	Sensitivity	PPV	NPV	OR*
Death						
PreOpNet (CSMC)	0·83 (0·79–0·87)	0·87 (0·86–0·88)	0·57 (0·48–0·68)	0·06 (0·05–0·08)	0·99 (0·99–0·99)	9·17 (5·85–13·82)
PreOpNet (CUMC)	0·79 (0·75–0·83)	0·86 (0·80–0·92)	0·49 (0·48–0·50)	0·03 (0·02–0·03)	1·00 (0·99–1·00)	6·20 (3·87–10·41)
PreOpNet (SHC)	0·75 (0·74–0·76)	0·93 (0·91–0·94)	0·32 (0·32–0·32)	0·02 (0·02–0·02)	1·00 (1·00–1·00)	5·88 (5·00–7·00)
RCRI	0·67 (0·61–0·72)	0·94 (0·93–0·94)	0·12 (0·05–0·19)	0·03 (0·01–0·05)	0·99 (0·98–0·99)	2·08 (0·77–3·50)
MACE						
PreOpNet (CSMC)	0·77 (0·73–0·80)	0·88 (0·88–0·89)	0·41 (0·33–0·49)	0·08 (0·06–0·10)	0·98 (0·98–0·99)	5·38 (3·75–7·49)
RCRI	0·63 (0·59–0·68)	0·94 (0·93–0·94)	0·10 (0·05–0·15)	0·04 (0·02–0·06)	0·98 (0·97–0·98)	1·67 (0·77–2·68)

CSMC denotes results for the internal held-out test dataset from the derivation cohort, and SHC and CUMC denote results for the external validation cohorts. AUC=area under the receiver operating characteristic curve. CSMC=Cedars-Sinai Medical Center. CUMC=Columbia University Medical Center. MACE=major adverse cardiovascular events. NPV=negative predictive value. OR=odds ratio. PPV=positive predictive value. RCRI=Revised Cardiovascular Risk Index. SHC=Stanford Healthcare. *Unadjusted ORs for the outcome of interest (death or MACE) were calculated for patients in the 85th percentile or higher of risk compared with those in the lower percentiles of risk, as determined by either PreOpNet or the RCRI (corresponding to an RCRI score ≥2).

Table 2: **Discrimination of postoperative outcomes by the PreOpNet or RCRI algorithms**

Discussion

In a large cohort of patients undergoing inpatient procedures, a deep-learning algorithm using the waveforms of a single preoperative 12-lead ECG identified risk for postoperative death for cardiac surgeries, non-cardiac surgeries, and catheterisation laboratory or endoscopy suite interventions. Compared with a widely used standard perioperative risk assessment tool and alternative ECG assessment tools, PreOpNet was able to more effectively identify high-risk patients who went on to experience postoperative mortality. Furthermore, the accuracy of PreOpNet for discriminating postoperative mortality was re-affirmed in two external health-care system cohorts with diverse patient populations. Additionally, in secondary analyses, PreOpNet identified high-risk patients who went on to experience MACE within the internal test dataset in the CSMC patient population. To our knowledge, PreOpNet is the first deep-learning architecture designed to aid clinicians in discriminating postoperative outcomes.

Over the past several years, the 12-lead ECG has been the subject of deep-learning algorithm development for potential clinical applications, with promising results. Previously applied to improve the detection of occult cardiovascular traits (eg, atrial fibrillation, hypertrophic cardiomyopathy, myocardial infarction, cardiac amyloidosis, and ventricular dysfunction), recent deep-learning ECG models have been shown to also identify non-cardiovascular-specific traits, such as liver disease,[30] anaemia,[26] age,[27] and long-term mortality.[23] These latter results highlight the potential of deep-learning methods to extract broad as well as specific novel information from ECG waveforms. Therefore, to optimise performance for the task of preoperative risk assessment, our algorithm used readily available ECG waveform data collected from large and diverse real-world patient cohorts. The resulting PreOpNet algorithm demonstrated ability not only to discriminate postoperative mortality, but to do so while outperforming a conventional clinical risk score and human-interpretable features from the ECG. As indicated in other clinical fields in which AI has leveraged latent features from a medical image to refine diagnosis or prognosis,[20] our results indicate the potential value of the PreOpNet algorithm to augment clinical decision making for preoperative risk assessment.

Several limitations of this preliminary study merit consideration. Many ambulatory procedures for clinically assessed low-risk patients do not involve acquiring a preoperative ECG. Therefore, this algorithm might not be applicable to such low-risk patients. Additionally, RCRI is most applicable and designed to be evaluated in patients undergoing non-cardiac surgery, so the most direct comparison is in this setting with an AUC of 0·83 for PreOpNet versus 0·70 for RCRI. Other comparisons only accentuate the difference in performance. In addition, all analyses were performed on retrospective cohorts. For each site, investigators at that site performed their own cohort selection as well as running inference, with only code and model weights shared across institutions. In our analyses, we noticed decreases in model performance in external validation, probably due to changes in cohort make-up and procedural definition. For example, data for the CUMC cohort were only collected from 3 months at the beginning of the COVID-19 pandemic and the SHC ECG data were extracted from a separate ECG data storage system, resulting in differences in cohort selection that could change model performance. Additional prospective validation studies are needed in large and diverse external cohorts—particularly of the exploratory, secondary MACE endpoint—to precisely evaluate PreOpNet's performance in discriminating events. Notwithstanding these limitations, the current study offers several strengths including the ability to leverage internal training, validation, and test datasets within a large derivation cohort of patients undergoing inpatient procedures over a decade. The algorithm was also able to be externally validated for post-procedural mortality in two large, diverse medical centres.

In summary, our findings demonstrate how a novel deep-learning algorithm, applied to a single preoperative ECG, can improve discrimination of postoperative adverse outcomes while running efficiently on a standard clinical workstation. Recognising that clinicians have limited time for making clinical assessments and decisions around potential post-procedural outcomes, conventional risk calculators using easily accessible information have been recommended by professional society practice guidelines to aid in perioperative risk stratification.[21,22] The opportunity to implement potentially more informative and easier-to-use prediction algorithms, in a manner that integrates with existing clinical workflows, offers a potential path towards improving postoperative outcomes. These promising results warrant further studies to establish the prospective validity of deep-learning algorithms for prognosticating post-procedural risk.

Contributors

DO, JWH, NRS, JT, JE, RKS, NY, PB, PE, JHC, JET, and MJ retrieved and quality-controlled all ECG data, merged electronic health record data, and verified the raw data. DO, BH, GD, and JT developed and trained the deep-learning algorithms. JWH, PE, TP, JET, JHC, JT, AP, and JYZ were responsible for external validation. MP, NY, MN, SSC, SC, and CMA performed clinical evaluation of model performance. DO, NRS, JT, SC, and CMA wrote the manuscript with critical review and feedback by all authors. DO, BC, and NRC performed the statistical analyses. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

DO reports support from the National Institute of Health (NIH; NHLBI R00HL157421) and Alexion, and consulting or honoraria for lectures from EchoIQ, Ultromics, Pfizer, InVision, the Korean Society of Echo, and the Japanese Society of Echo. JWH reports funding from the National Science Foundation (DGE-1656518). PE reports research support from the NIH (T32HL007854-21) and serves on the American College of Cardiology Innovation Council. BC reports consulting from Alnylam, Cardurion, Corvia, CVRX, Cytokinetics, Intellia, and Rocket. TP reports stock in Abbot and Baxter, and research support from the

American Heart Association, Eido, Pfizer, Edwards, and the New York Academy of Medicine. JHC reports research funding from NIH/National Institute of Aging (AI17812101), NIH/National Institute on Drug Abuse (UG1DA015815-CTN-0136), AIMI-HAI Partnership, Doris Duke Charitable Foundation (20211260), Google (SPO136094), American Heart Association, and Clinical and Translational Sciences Awards Program by the National Center for Advancing Translational Sciences (R56LM013365); is a co-founder of Reaction Explorer; and receives consulting fees from Sutton Pierce and Younker Hyde MacFarlane. MP reports research support from the NIH (HL136390), patents for ECG-based cardiac arrhythmia detection (WO2014205310A3), and consulting fees from Apple, Biotronik, Boston Scientific, QALY, Johnson and Johnson, and Bristol Myers Squibb. CMA reports support from NIH (HL165840), consulting fees from Medtronic, Novartis, Illumina, and Medtronic, and participation in advisory boards for Boston Scientific, Medtronic, and Element Science; and is a trustee on the board for the Heart Rhythm Society. All other authors declare no competing interests.

Data sharing

All code and analytical methods applied for the algorithm and analyses are available at: https://github.com/ecg-net/PreOpNet. The patient data are not available given their potentially identifiable nature.

References

1 Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol* 2021; **18:** 465–78.

2 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:** 1043–49.

3 Lindenauer PK, Pekow P, Wang K, Mamidi DK, Gutierrez B, Benjamin EM. Perioperative beta-blocker therapy and mortality after major noncardiac surgery. *N Engl J Med* 2005; **353:** 349–61.

4 Fronczek J, Polok K, Devereaux PJ, et al. External validation of the Revised Cardiac Risk Index and National Surgical Quality Improvement Program Myocardial Infarction and Cardiac Arrest calculator in noncardiac vascular surgery. *Br J Anaesth* 2019; **123:** 421–29.

5 Davis C, Tait G, Carroll J, Wijeysundera DN, Beattie WS. The Revised Cardiac Risk Index in the new millennium: a single-centre prospective cohort re-evaluation of the original variables in 9519 consecutive elective surgical patients. *Can J Anaesth* 2013; **60:** 855–63.

6 Ford MK, Beattie WS, Wijeysundera DN. Systematic review: prediction of perioperative cardiac complications and mortality by the revised cardiac risk index. *Ann Intern Med* 2010; **152:** 26–35.

7 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:** 381–87.

8 Glance LG, Faden E, Dutton RP, et al. Impact of the choice of risk model for identifying low-risk patients using the 2014 American College of Cardiology/American Heart Association perioperative guidelines. *Anesthesiology* 2018; **129:** 889–900.

9 Sheth T, Chan M, Butler C, et al. Prognostic capabilities of coronary computed tomographic angiography before non-cardiac surgery: prospective cohort study. *BMJ* 2015; **350:** h1907.

10 Rao JY, Yeriswamy MC, Santhosh MJ, et al. A look into Lee's score: peri-operative cardiovascular risk assessment in non-cardiac surgeries—usefulness of revised cardiac risk index. *Indian Heart J* 2012; **64:** 134–38.

11 Park SJ, Choi JH, Cho SJ, et al. Comparison of transthoracic echocardiography with N-terminal pro-brain natriuretic peptide as a tool for risk stratification of patients undergoing major noncardiac surgery. *Korean Circ J* 2011; **41:** 505–11.

12 Choi JH, Cho DK, Song YB, et al. Preoperative NT-proBNP and CRP predict perioperative major cardiovascular events in non-cardiac surgery. *Heart* 2010; **96:** 56–62.

13 Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg* 2013; **217:** 833–42.

14 Glance LG, Lustik SJ, Hannan EL, et al. The Surgical Mortality Probability Model: derivation and validation of a simple risk prediction rule for noncardiac surgery. *Ann Surg* 2012; **255:** 696–702.

15 Machado MN, Rodrigues FB, Nakazone MA, et al. Prediction of death after noncardiac surgery: potential advantage of using high-sensitivity troponin T as a continuous variable. *J Am Heart Assoc* 2021; **10:** e018008.

16 Marston N, Brenes J, Garcia S, et al. Peak postoperative troponin levels outperform preoperative cardiac risk indices as predictors of long-term mortality after vascular surgery: troponins and postoperative outcomes. *J Crit Care* 2012; **27:** 66–72.

17 Devereaux PJ, Chan MT, Alonso-Coello P, et al. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. *JAMA* 2012; **307:** 2295–304.

18 Ouyang D, He B, Ghorbani A, et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* 2020; **580:** 252–56.

19 McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020; **577:** 89–94.

20 Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* 2018; **2:** 158–64.

21 Apfelbaum JL, Connis RT, Nickinovich DG, et al. Practice advisory for preanesthesia evaluation: an updated report by the American Society of Anesthesiologists Task Force on Preanesthesia Evaluation. *Anesthesiology* 2012; **116:** 522–38.

22 Fleisher LA, Fleischmann KE, Auerbach AD, et al. 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *J Am Coll Cardiol* 2014; **64:** e77–137.

23 Raghunath S, Ulloa Cerna AE, Jing L, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat Med* 2020; **26:** 886–91.

24 Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med* 2019; **25:** 70–74.

25 Galloway CD, Valys AV, Shreibati JB, et al. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. *JAMA Cardiol* 2019; **4:** 428–36.

26 Kwon J-M, Cho Y, Jeon K-H, et al. A deep learning algorithm to detect anaemia with ECGs: a retrospective, multicentre study. *Lancet Digit Health* 2020; **2:** e358–67.

27 Attia ZI, Friedman PA, Noseworthy PA, et al. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circ Arrhythm Electrophysiol* 2019; **12:** e007284.

28 Suchard MA, Schuemie MJ, Krumholz HM, et al. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *Lancet* 2019; **394:** 1816–26.

29 Singh S, Fouayzi H, Anzuoni K, et al. Diagnostic algorithms for cardiovascular death in administrative claims databases: a systematic review. *Drug Saf* 2019; **42:** 515–27.

30 Fleisher LA, Fleischmann KE, Auerbach AD, et al. 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* 2014; **130:** e278–333.