

Assessment of Functional Capacity Before Major Non-Cardiac Surgery

Duminda N Wijesundera, PhD; Rupert M Pearce, MD; Mark A Shulman, MPH; Tom E F Abbott, PhD; Elizabeth Torres, BMSc; Althea Ambosta, MA; Bernard L Croal, MD; John T Granton, MD; Kevin E Thorpe, MMath; Michael P W Grocott, MD; Catherine Farrington, BAppSc; Paul S Myles, DSc; and Brian H Cuthbertson, MD; on behalf of the METS Study Investigators

Author Affiliations

From the Li Ka Shing Knowledge Institute (DNW, KET), and Applied Health Research Centre (ET, AA, KET), St. Michael's Hospital; Department of Anesthesia and Pain Management (DNW), and Department of Medicine (JTG), University Health Network; Department of Anesthesia (DNW, BHC), Institute of Health Policy Management and Evaluation (DNW, BHC), Department of Medicine (JTG), and Dalla Lana School of Public Health (KET), University of Toronto; Department of Medicine, Sinai Health System (JTG); Department of Critical Care Medicine, Sunnybrook Health Sciences Centre (BHC) – all in Toronto, Ontario, Canada; Queen Mary University of London, London (RMP, TEFA), NHS Grampian, Aberdeen (BLC), University Hospital Southampton, Southampton (MPWG), and University of Southampton, Southampton (MPWG) – all in the United Kingdom; and the Alfred Hospital (MAS, CF, PSM) and Monash University (MAS, CF, PSM) – all in Melbourne, Victoria, Australia.

Correspondence

Brian H. Cuthbertson, MB ChB MD

Department of Critical Care Medicine
Sunnybrook Health Sciences Centre
2075 Bayview Avenue, Toronto, Ontario
Canada M4N 3M5
Email: brian.cuthbertson@sunnybrook.ca

Word count: 3699

RESEARCH IN CONTEXT

Evidence Before This Study: Evaluation of cardiopulmonary fitness, or functional capacity, is a highly-emphasised component of risk assessment before major non-cardiac surgery. This evaluation typically involves “subjective assessment”, where doctors interview patients and make a subjective judgement of their fitness. To assess the validity of this commonly used measure of functional capacity, we used the terms [(“prediction” OR “preoperative evaluation” OR “risk prediction”) AND (“surgery” AND “complications”) AND (“exercise capacity” OR “activities of daily living” OR “functional capacity”)] to search the Pubmed database for relevant studies published before 31 December 2017. This search was supplemented with hand-searches of reference lists from relevant reviews and practice guidelines. Prior research was limited to single-centre studies with small sample sizes or a high risk of bias. In these studies, subjective assessment showed poor agreement with validated questionnaires, and an inconsistent association with postoperative complications.

Added Value of This Study: The Measurement of Exercise Tolerance before Surgery (METS) was a multicentre prospective cohort study in major elective non-cardiac surgery that compared the prognostic accuracy of subjective assessment against three alternatives, namely the Duke Activity Status Index (DASI) questionnaire, cardiopulmonary exercise testing (CPET) to measure peak oxygen consumption (VO_2 peak), and serum N-terminal pro-B-type natriuretic peptide (NT pro-BNP) concentrations. In a sample of 1401 adult participants at 25 hospitals, DASI predicted 30-day death or myocardial infarction, and 30-day death or myocardial injury; NT pro-BNP predicted 30-day death or myocardial injury, and one-year death; and VO_2 peak predicted complications. Subjective assessment did not predict any outcomes.

Implications of All the Available Evidence: Subjective assessment of functional capacity should not be used for preoperative risk evaluation. This commonly employed practice does not accurately identify patients with poor fitness or those at elevated risk for postoperative morbidity and mortality. As alternatives, clinicians can consider more objective measures, such as DASI questionnaires and NT pro-BNP testing for evaluating perioperative cardiac risk, and perhaps CPET to predict complications after major elective non-cardiac surgery.

SUMMARY

Background: Functional capacity is an emphasized component of risk assessment for major surgery. Doctors' clinical assessment of patients' functional capacity has uncertain accuracy.

Methods: This international prospective cohort study at 25 hospitals recruited adults (≥ 40 y) scheduled for major non-cardiac surgery. Functional capacity was subjectively assessed as <4 (poor), 4 to 10 (moderate), or >10 (good) metabolic equivalents. Participants completed the Duke Activity Status Index (DASI) questionnaire, cardiopulmonary exercise testing (CPET) to measure peak oxygen consumption (VO_2 peak), and blood testing for N-terminal pro-B-type natriuretic peptide (NT pro-BNP). Main outcomes were 30-day death or myocardial infarction (primary outcome), one-year death (secondary outcome), moderate-or-severe postoperative complications, and 30-day death or myocardial injury. Prognostic accuracy was evaluated using logistic regression, receiver-operating-characteristic curves and net risk reclassification.

Findings: Of 1401 participants, 2.0% experienced 30-day death or myocardial infarction, 2.7% experienced one-year death, 13.9% experienced complications, and 12.6% experienced 30-day death or myocardial injury. Subjective assessment had 19.2% (95% CI, 14.2% – 25.0%) sensitivity and 94.7% (93.2% – 95.9%) specificity for identifying the inability to attain 4 metabolic equivalents during CPET. Subjective assessment was not associated with outcomes. DASI was associated with the primary outcome ($p=0.03$), and 30-day death or myocardial injury ($p=0.05$). NT pro-BNP was associated with 30-day death or myocardial injury ($p=0.003$), and one-year death ($p=0.001$). VO_2 peak was associated with complications ($p=0.007$).

Interpretation: Subjectively assessed functional capacity should not be used for preoperative risk evaluation. Clinicians should instead consider DASI or NT pro-BNP for cardiac risk evaluation, and CPET for predicting postoperative complications.

Funding: Canadian Institutes of Health Research; Heart and Stroke Foundation of Canada; Ontario Ministry of Health and Long-Term Care; Ontario Ministry of Research, Innovation and Science; UK National Institute of Academic Anaesthesia; UK Clinical Research Collaboration; Australian and New Zealand College of Anaesthetists; Monash University.

INTRODUCTION

International clinical practice guidelines emphasise evaluation of preoperative cardiopulmonary fitness, or functional capacity, as a critical component of the assessment of patients' risks for major morbidity and mortality after surgery.^{1,2} For example, the American College of Cardiology and American Heart Association recommend that patients proceed directly to elective intermediate and major non-cardiac surgery if they are capable of more than four metabolic equivalents of activity without symptoms.¹ The current usual standard of care for assessing preoperative functional capacity involves doctors evaluating patients and then making subjective estimates of their fitness (hereafter referred to as 'subjective assessment'). While easily implementable into clinical practice, subjective assessment has limitations, including poor agreement with validated measures of functional capacity,³ and relatively poor accuracy when predicting postoperative death or complications.^{4,5} These limitations point to the need for better alternatives to assess preoperative functional capacity.

Possible options are cardiopulmonary exercise testing (CPET), which has been described as a "gold standard" non-invasive assessment of exercise tolerance,⁶ and the Duke Activity Status Index (DASI),⁷ which is a standardised questionnaire correlated with gold-standard measures of functional capacity. Additionally, while no blood test can directly measure functional capacity, N-terminal pro-B-type natriuretic peptide (NT pro-BNP) concentrations might indirectly fulfil this role by serving as an integrated marker of cardiac dysfunction.⁸ We therefore conducted the Measurement of Exercise Tolerance before Surgery (METS) study to compare preoperative

subjective assessment, CPET, DASI, and NT pro-BNP for predicting death or complications after major elective non-cardiac surgery.

METHODS

Study Design

The METS study was a multicentre prospective cohort study conducted in Canada, the United Kingdom, Australia, and New Zealand. Its objectives, design, and methods have been previously reported.⁹ All participants provided written informed consent, and each centre obtained research ethics board approval before commencing recruitment. Details of the methods are presented in the Online Appendix.

Participants and Procedures

Participants were aged 40 years or older, scheduled for elective non-cardiac surgery under general and/or regional anaesthesia with a minimum of an overnight hospital stay, and deemed to have one or more risk factors for cardiac complications or coronary artery disease (Tables S1 and S2 – Online Appendix). During the period from recruitment to one day before surgery, participants underwent symptom-limited incremental CPET on a computer-controlled, electromagnetically braked cycle ergometer using a standardised protocol (Online Appendix).⁹ This assessment usually occurred during a separate hospital visit after the date of recruitment. Based on a protocolised evaluation of the plotted CPET data, trained investigators at each centre determined both the peak oxygen consumption (VO₂ peak) and anaerobic threshold (AT).

Participants also underwent three other preoperative assessments of functional capacity. First, responsible anaesthesiologists in the preoperative assessment clinic (on date of recruitment) or operating theatre (on day of surgery) were asked to make a subjective judgment of participants'

functional capacity after conducting their usual preoperative history. Subjectively assessed functional capacity was classified as ‘poor’ (<4 metabolic equivalents), ‘moderate’ (4 to 10 metabolic equivalents), and ‘good’ (>10 metabolic equivalents). The ‘poor’ category included cases where anaesthesiologists remained uncertain about patients’ functional capacity after conducting their usual preoperative history, typically due to pre-existing conditions such as arthritis or peripheral arterial disease.¹ Second, participants completed the DASI questionnaire on the date of recruitment (Online Appendix – Table S3). Third, blood samples were drawn at any point between recruitment and surgery to measure serum NT pro-BNP concentrations. These samples were stored at -70°C to -80°C in each study site, and then analysed at the Aberdeen Royal Infirmary (Aberdeen, UK) using the Siemens Vista™ immunoassay analyser (Siemens Healthcare Diagnostics Ltd, Frimley, UK). Participants, healthcare providers and outcome adjudicators were blinded to CPET and NT pro-BNP results, while healthcare providers and outcome adjudicators were blinded to DASI scores. The exceptions were cases of myocardial ischaemia or significant new arrhythmias during CPET, in which case these specific findings, but not VO₂ peak or AT, were revealed to healthcare providers.

After surgery, participants underwent daily electrocardiograms and blood sampling to measure troponin and creatinine concentrations, until the third postoperative day or hospital discharge (whichever came first). Research personnel followed participants daily throughout their hospital stay to ascertain the presence of specific complications (Online Appendix – Table S4). The severity of complications was further categorised as mild, moderate, severe or fatal using a modified Clavien-Dindo classification scheme.^{10,11} After hospital discharge, participants were

contacted at 30 days and one year after surgery to ascertain vital status. Details of the follow-up process are presented in the Online Appendix.

Study Outcomes

The primary outcome was death or myocardial infarction within 30 days after surgery, while the secondary outcome was death within one year after surgery. Myocardial infarction was diagnosed by an adjudication committee that used the Third Universal Definition of Myocardial Infarction while remaining blinded to CPET, DASI, and NT pro-BNP results.¹² Other outcomes of interest were death or myocardial injury within 30 days after surgery, and moderate-or-severe (including fatal) complications during the index hospitalisation. Myocardial injury was defined as postoperative troponin concentrations exceeding both the 99th percentile of the normal reference population, and the threshold at which the assay coefficient of variation was 10%. Moderate-or-severe complications were included as an endpoint because these events have been associated with poor preoperative functional capacity, especially as measured objectively by CPET.^{13,14}

Statistical Analyses

In the primary analysis, all participants who undertook both CPET and surgery were included, and CPET performance was characterised by VO₂ peak. For each outcome of interest, we built separate nested logistic regression models that sequentially included baseline clinical characteristics, and then followed by the exposure of interest (i.e. subjective assessment, VO₂ peak, AT, DASI, or NT pro-BNP). We modelled NT pro-BNP concentrations using a logarithmic transformation to reduce the potential influence of extreme values within its highly-

skewed distribution. The statistical significance of prognostic information from additional predictors was based on the increase in log likelihood of the “larger” model. For the models predicting the primary and secondary outcomes, the baseline variable was the validated Revised Cardiac Risk Index (RCRI) score.^{15,16} In the model predicting 30-day death or myocardial injury, the baseline variables were age, sex, and RCRI score. The baseline variables in the model predicting moderate-or-severe complications were age, sex, and high-risk surgery, which was defined as intra-peritoneal, intra-thoracic, or supra-inguinal vascular procedures.¹⁵ These covariates were selected *a priori* based on prior evidence, their inclusion in guideline-recommended assessment algorithms,^{1,17} and the need to prevent model overfitting.¹⁸ Additionally, the covariates mirror clinically-sensible factors typically considered during preoperative evaluation. We calculated the area under the receiver-operating-characteristic (ROC) curve of models with successively more predictors, as well as models with only the individual exposure of interest (e.g., VO₂ peak). Prognostic information from these models was compared using the continuous net reclassification improvement (NRI) statistic and area under the ROC curve.^{19,20}

The sample size calculation was based on comparing the area under the curve (AUC) of ROC curves for VO₂ peak versus subjective assessment with respect to predicting 30-day myocardial infarction or death. During the design of the METS study, a required sample size of 1180 was initially calculated based on the underlying assumptions of an outcome event rate of 8%, correlation of 0·5 between VO₂ peak and subjective assessment,⁷ AUC of 0·65 for subjective assessment, AUC of 0·75 for VO₂ peak, and 90% power to detect this difference in AUC values (2-sided alpha of 0·05). To account for 10% of participants not undertaking CPET or surgery, we

aimed to recruit 1312 participants. After recruiting about half of the original planned sample size, this sample size calculation was re-evaluated based on two factors identified in the accumulating study data. First, we found that about 20% of participants did not either undertake CPET or undergo their planned surgeries. Second, the event rate for the primary outcome was instead projected to be approximately 5%. The overall sample size was therefore increased to 1723 participants to account for up to 20% of recruited individuals not being eligible for the primary analysis, and a primary outcome event rate of 5%, while retaining the power of 80%. Importantly, we remained blinded to all data on the principal exposures (i.e., CPET results, DASI scores, NT pro-BNP concentrations) during this sample size re-estimation.

Analyses were conducted using the R statistical language (Version 3.4.0), statistical significance was defined by a two-tailed p-value <0.05 and no adjustments were made for multiple comparisons.²¹ Since missing data were very uncommon ($<4\%$), a complete case analysis was performed.²² Additionally, these missing data pertained to baseline information that was likely missing completely at random.²³

RESULTS

Participant Characteristics

From March 2013 to March 2016, 1741 patients were recruited at 25 hospitals, with 1401 (80·5%) undergoing CPET and surgery (Figure 1). Of these 1401 participants in the primary cohort, 1399 (99·9%) completed 30-day follow-up and 1378 (98·4%) completed one-year follow-up. The characteristics of the cohort are presented in Table 1.²⁴ Their mean age was 64 years, 39% were female, and 91% were classified as American Society of Anesthesiologists Physical Status (ASA-PS) 2 or 3. Most participants underwent major abdominal, pelvic, or orthopaedic procedures. Arterial lines were inserted in 49% of participants, while central venous catheters were inserted in 15%. After surgery, 24% of participants were admitted to critical care or monitored bed units.

Assessments of Preoperative Functional Capacity

The CPET exercise protocol was terminated early in 11% of participants, with the most common reasons being inability to pedal, fatigue, or a safety-based indication (Table 2). A VO₂ peak was measurable in 1356 (97%) participants, with the mean being 19·2 mL/kg/min (SD 6·5). By comparison, 1275 (91%) participants had a measurable AT, with a mean of 12·6 mL/kg/min (SD 4·1). Adverse events during CPET were experienced by 8% of participants (Table 2), with 27 cases meeting criteria for unblinding. The clinical sequelae of these events are presented in Table S5 (Online Appendix).

In addition, 1351 (96·4%) participants underwent subjective assessment of functional capacity, 1396 (99·6%) completed DASI questionnaires, and 1347 (96·1%) had NT pro-BNP concentrations measured (Table S6 – Online Appendix). About 8% of participants (n=107) were subjectively judged to have poor preoperative functional capacity, 16% (n=230) had a VO₂ peak <14 mL/kg/min (equivalent to <4 metabolic equivalents), and 30% (n=426) had an AT below the suggested high-risk threshold of 11 mL/kg/min.²⁵ The characteristics of participants within strata defined by VO₂ peak is presented in Table S7 (Online Appendix). A subjective assessment of poor functional capacity had a sensitivity of 19·2% (95% CI, 14·2% – 25·0%) for identifying VO₂ peak <14 mL/kg/min, while its specificity was 94·7% (93·2% – 95·9%).

When stratified by subjectively assessed functional capacity, VO₂ peak and DASI values were generally lower in individuals judged to have poor functional capacity, but there was considerable within-stratum variation (Figure 2: Panels A and B). VO₂ peak (Figure 2: Panels C and D) was positively correlated with DASI scores (Spearman rho 0·43, p<0·0001), and negatively correlated with NT pro-BNP concentrations (Spearman rho -0·21, p<0·0001). There was also a negative correlation (Figure S1 – Online Appendix) between DASI scores and NT pro-BNP concentrations (Spearman rho -0·25, p<0·0001).

Postoperative Outcomes

After surgery, 194 (13·9%) participants suffered in-hospital moderate-or-severe complications. By 30 days after surgery, 5 (0·4%) participants had died, 24 (1·7%) experienced myocardial infarction, 28 (2·0%) experienced the primary outcome of death or myocardial infarction, and 176 (12·6%) experienced death or myocardial injury. By one year after surgery, 38 (2·7%)

participants had died. Among moderate-or-severe complications, the more frequent events were respiratory failure, pneumonia, surgical site infection, re-operation, and unexpected critical care unit admission (Table S9 – Online Appendix).

Prognostic Performance of Preoperative Functional Capacity Assessments

Subjective assessment ratings, VO₂ peak, AT, DASI scores and NT pro-BNP concentrations within strata defined by the four outcomes of interest are presented in Tables S8 and S9 (Online Appendix). Subjectively assessed preoperative functional capacity had no statistically significant adjusted association with the four main study outcomes (Table S10 – Online Appendix).

The adjusted associations of VO₂ peak, AT, DASI scores and NT pro-BNP concentrations with the main outcomes of interest, as well as continuous NRI statistics describing their incremental predictive performance, are presented in Table 3. For VO₂ peak, a statistically significant adjusted association and significant risk reclassification was observed only with respect to moderate-or-severe complications. Anaerobic threshold showed no statistically significant association or risk reclassification with the main outcomes. DASI scores showed statistically significant adjusted associations with (i) 30-day death or myocardial infarction and (ii) 30-day death or myocardial injury. Additionally, it showed significant risk reclassification with 30-day death or myocardial injury. Conversely, NT pro-BNP concentrations showed statistically significant adjusted associations and significant risk reclassification with both (i) 1-year death and (ii) 30-day death or myocardial injury.

Post-Hoc Subset Analysis in Body Cavity Surgery

To address whether preoperative functional capacity might have better prognostic value in more invasive surgical procedures, a *post-hoc* subset analysis was performed in body cavity surgery, which was defined as intrathoracic, intraperitoneal, retroperitoneal or pelvic (i.e., urologic or gynaecologic) procedures. The characteristics and outcomes of the 912 participants in this subset are presented in Table S11 (Online Appendix). When the main study analyses were repeated in this subset, the results remained qualitatively unchanged (Table S12 – Online Appendix).

DISCUSSION

In this international prospective cohort study of non-cardiac surgery patients, preoperative subjective assessment neither accurately identified patients with poor cardiopulmonary fitness, nor predicted postoperative morbidity and mortality. The DASI questionnaire improved prediction of 30-day myocardial infarction or death, and 30-day myocardial injury or death; while NT pro-BNP concentrations improved prediction of 30-day myocardial injury or death, and 1-year death. Formal evaluation of cardiopulmonary fitness, based on VO₂ peak during CPET, improved prediction of moderate-or-severe postoperative complications.

Interpretation and Comparison with Prior Research

In the METS study, subjective assessment of preoperative functional capacity consistently demonstrated poor performance. While it had construct validity, in that VO₂ peak was generally lower in patients judged to be less fit, subjective assessment correctly identified only 16% of patients who achieved a VO₂ peak less than 14 mL/kg/min, which is consistent with less than four metabolic equivalents. Further, subjective assessment did not predict postoperative myocardial infarction, myocardial injury or complications, confirming a prior single-centre retrospective cohort study that relied on administrative database for outcome ascertainment.⁵ Based on these findings, subjective assessment should not be used to assess patients' risks of major postoperative cardiac complications.

Notably, more objective evaluation of cardiopulmonary fitness with CPET did not improve most aspects of preoperative risk assessment. Consistent with prior evidence,^{13,14} VO₂ peak measured

during CPET was predictive of postoperative complications; however, most of these events were pulmonary complications, surgical site infections, unexpected critical care unit admissions, and re-operations. In contrast, VO₂ peak and AT were not associated with postoperative myocardial infarction or myocardial injury, somewhat contradicting the emphasis of practice guidelines on functional capacity for preoperative cardiac risk evaluation.^{1,2} These findings occurred within the context of our study addressing several important limitations in the current evidence base,²⁶ in that it blinded CPET results (unlike most previous studies), and implemented standardised outcome surveillance in a large generalisable multicentre sample. There are several possible explanations for our results. First, prior evidence supporting a link between fitness and perioperative cardiac risk has limitations. For example, several studies were conducted more than 30 years ago, and have limited generalisability to contemporary patients and surgeries.^{27,28} Other studies observed few outcome events or associations of only weak magnitudes.^{29,30} Second, a low VO₂ peak or AT might not be the ideal CPET-based indicator of the underlying causal mechanisms for perioperative myocardial infarction. It is possible that other metrics, such as an exaggerated exercise-mediated heart rate response,³¹ are better indicators of perioperative cardiac risk.

The METS study found the DASI questionnaire to have construct validity as a measure of functional capacity in surgical patients, consistent with prior research.³² Importantly, it also improved prediction of postoperative myocardial infarction and myocardial injury. Our findings confirm non-operative data indicating enhanced risk prediction using this questionnaire,^{33,34} support guideline suggestions for using objective scales to evaluate functional capacity,¹ and point to opportunities for straightforward improvements in clinical practice. Specifically, the

simple DASI questionnaire can be easily implemented into most perioperative practice settings, although further work is needed to define optimal risk-specific thresholds in DASI scores, and develop reliable non-English versions of the questionnaire.³⁵ An important area of residual uncertainty is why DASI scores were associated postoperative cardiac events, yet VO₂ peak was not. Given the only moderate correlation between DASI scores and VO₂ peak, a possible explanation is that DASI measures also somewhat different constructs, such as musculoskeletal strength, frailty, and self-imposed physical limitations.³⁶

Confirming a previous individual patient data meta-analysis,³⁷ elevated preoperative NT pro-BNP concentrations were associated with increased risks of postoperative 30-day death or myocardial injury in the METS study cohort; additionally, elevated concentrations also predicted increased one-year mortality. These findings support recommendations in recent practice guidelines to incorporate natriuretic peptide testing into preoperative risk assessment strategies.¹⁷ Notably, we found only slight-to-fair correlation between NT pro-BNP concentrations and measures of exercise capacity (i.e., VO₂ peak, DASI). This low correlation suggests that NT pro-BNP measures a construct distinct from exercise capacity, and raises that possibility for enhancing preoperative assessment by combining measures of functional capacity and NT pro-BNP in future risk prediction models.

Limitations

Our study has several limitations. First, despite increasing the original projected sample size, the event rates for the primary and secondary outcomes were less than anticipated. To some extent, the risks of 30-day death (i.e., 0.4%) and 30-day death and or myocardial infarction (i.e., 2.0%)

in the METS study are representative of contemporary *elective* major non-cardiac surgery in high-income countries, as suggested by several studies published after the design of our study. For example, in the International Surgical Outcomes (ISOS) prospective cohort study of 44,814 adults having elective inpatient surgery across 27 high and middle income countries, the risk of in-hospital 30-day death was 0.5%.¹¹ Similarly, in the Vascular Events In Noncardiac Surgery Patients Cohort Evaluation (VISION) prospective cohort study of 15,133 patients having inpatient non-cardiac surgery in eight high and middle income countries, the risk of 30-day mortality after elective surgery was 1.2%. In this same VISION cohort, the overall risk of myocardial infarction was about 3.3%.³⁸ Since emergency surgery constituted 14% of the cohort, and is associated with a three-fold higher risk of myocardial infarction,³⁹ the risk of myocardial infarction after elective surgery in the VISION study was likely about 2.6%. Thus, the event rates in the METS study are consistent with other contemporary major non-cardiac surgery samples. Nonetheless, to help address this reduced statistical power, we analysed the association of the exposures of interest with two more frequent outcomes: myocardial injury and moderate-or-severe postoperative complications. Myocardial injury and postoperative complications are clinically and prognostically important outcomes.³⁹⁻⁴² Our general findings with respect to prediction of 30-day myocardial infarction or death were qualitatively unchanged in the complementary analysis pertaining to 30-day myocardial injury or death.

Second, despite significant efforts by research personnel and study investigators, the consent rate among otherwise eligible patients in the METS study was 27%. Nonetheless, this consent rate is somewhat unsurprising when viewed from the perspective of the study setting (i.e., anxious patients awaiting major surgery within a short time frame) and procedures (i.e., strenuous

exercise testing solely for research purposes). It is also consistent with several large contemporary prospective studies in surgical patients that had arguably more straightforward study procedures. For example, consent rates were 30% in the Perioperative Ischemic Evaluation 2 (POISE-2) trial of aspirin and clonidine in non-cardiac surgery,⁴³ and 36% in the Aspirin and Tranexamic Acid for Coronary Artery Surgery (ATACAS) trial in cardiac surgery.⁴⁴

Third, the primary analyses relied on VO₂ peak and AT determined by trained investigators at each individual centre. It is possible that central adjudication of CPET results might have led to different determinations of VO₂ peak and AT. Nonetheless, given that the METS study was designed to be pragmatic and generalisable, our main analyses better represent the prognostic accuracy of VO₂ peak or AT in real world clinical practice. Fourth, the preoperative predictive models in this study had generally low-to-moderate discrimination, with AUC values of 0.74 or lower. This observation is, in part, explained by the lower-than-expected outcome event rate, which limited the number of covariates included in regression models. Nonetheless, the discrimination of these models is comparable to recent research, such as a prospective cohort study where the combination of RCRI score and preoperative coronary computed tomographic angiography had an AUC of 0.66 for predicting 30-day death or myocardial infarction.⁴⁵ Fifth, each hospital used its own preferred troponin assay to detect myocardial infarction or myocardial injury. This pragmatic approach is consistent with many contemporary multicentre perioperative studies.^{43,46} Furthermore, variation in troponin assays does not influence the prognostic importance of myocardial infarction or myocardial injury,^{39,41} and should not have biased the association between the study exposures and outcomes.

Future Directions

In combination with previous data, the results of the METS study now suggest that DASI scores and natriuretic peptides, such as NT pro-BNP, should supplant subjective assessment for the estimation of perioperative cardiac risk for major non-cardiac surgery. More research is required to define optimal thresholds for these measures, and determine how they should be utilised in combination with other prognostically important information, including alternative preoperative biomarkers (e.g., high-sensitivity troponin).⁴⁷ These other biomarkers might also help address, in part, the limitations of NT pro-BNP as a prognostic biomarker in the presence of obesity,⁴⁸ chronic renal kidney,⁴⁹ or heart failure with preserved ejection fraction.⁵⁰

Our findings also indicate that CPET, specifically VO₂ peak, can be used to identify patients at elevated risk for postoperative complications. Research remains needed to define optimal thresholds in VO₂ peak, determine the role of central expert adjudication to improving its prognostic accuracy, evaluate novel CPET-derived metrics of cardiopulmonary fitness (e.g., heart rate recovery), and assess possible inter-relationships of CPET-derived metrics with prognostically important comorbidities. For example, lower haemoglobin concentrations are associated with both elevated perioperative risk and reduced VO₂ peak.^{51,52}

Conclusions

Preoperative subjective assessment of functional capacity should not be used in clinical practice because it does not accurately identify patients with poor fitness or those at elevated risk for morbidity and mortality after major elective non-cardiac surgery. Clinicians can instead consider more objectives measures, such as DASI questionnaires and NT pro-BNP testing for evaluating

perioperative cardiac risk, and perhaps CPET to predict complications after major elective non-cardiac surgery.

ACKNOWLEDGMENTS

This study was supported by grants from the Canadian Institutes of Health Research; Heart and Stroke Foundation of Canada; Ontario Ministry of Health and Long-Term Care; Ontario Ministry of Research, Innovation and Science; United Kingdom (UK) National Institute of Academic Anaesthesia; UK Clinical Research Collaboration; Australian and New Zealand College of Anaesthetists; and Monash University (Melbourne, Victoria, Australia). These sponsors had no role in the design and conduct of the METS Study; collection, management, analysis and interpretation of the data; preparation, review or approval of this protocol paper; and decision to submit this manuscript for publication.

Dr Wijeysondera is supported by a New Investigator Award from the Canadian Institutes of Health Research. Dr Wijeysondera and Prof Cuthbertson are supported in part by Merit Awards from the Department of Anesthesia at the University of Toronto. Prof Pearce is a British Journal of Anaesthesia and Royal College of Anaesthetists Career Development Fellow, as well as a UK National Institute for Health Research Professor. Dr Abbott is a Medical Research Council and British Journal of Anaesthesia Clinical Research Training Fellow. Prof Grocott holds the British Oxygen Company Chair of Anaesthesia of the Royal College of Anaesthetists, which is awarded by the UK National Institute of Academic Anaesthesia. The authors thank the Li Ka Shing Knowledge Institute of St. Michael's Hospital (Toronto, Ontario, Canada) for generously supporting the costs of international trial insurance for this study. Finally, the authors thank all the participating patients and staff across the 25 study sites.

FIGURE LEGENDS

Figure 1: Participant screening, recruitment and follow-up in the METS study

Figure 2: Correlation between measures of preoperative functional capacity (Panels A to D)

Legend:

- A. Panel A is a boxplot presenting the distributions of VO₂ peak within strata defined by subjectively assessed preoperative functional capacity, which is categorised as poor (<4 metabolic equivalents), moderate (4-10 metabolic equivalents) or good (>10 metabolic equivalents). The horizontal line within each box denotes the median, while the top and bottom of each box indicate the interquartile range. Vertical lines at the top and bottom of each box extend to the 5th and 95th percentiles, while solid circles indicate outliers. Four metabolic equivalents correspond to a VO₂ peak of 14 mL/kg/min, while 10 metabolic equivalents correspond to a VO₂ peak of 35 mL/kg/min. The boxplot on the far right (denoted as 'NA') presents the VO₂ peak distribution among the 50 participants with missing subjective assessments. The wide-ranging distribution in this subgroup suggests that these values were missing completely at random.
- B. Panel B is a boxplot presenting the distributions of DASI scores within strata defined by subjectively assessed preoperative functional capacity. The minimum possible DASI score is 0, while 58.2 is maximum possible score. The boxplot on the far right (denoted as 'NA') presents the DASI score distribution among the 50 participants with missing subjective assessments. The wide-ranging distribution in this subgroup suggests that these values were missing completely at random.

- C. Panel C is a scatter plot presenting the association of VO₂ peak (y-axis) and DASI scores (x-axis). The Spearman correlation coefficient between DASI scores and VO₂ peak is 0·41 ($p<0\cdot001$). The blue line is a line of best fit (estimated using cubic regression splines), while the grey shaded zone represents its 95% confidence limits.
- D. Panel D is a scatter plot presenting the association of VO₂ peak (y-axis) and NT pro-BNP concentrations (x-axis on a log₁₀ scale). The Spearman correlation coefficient between NT pro-BNP and VO₂ peak is -0·21 ($p<0\cdot0001$). The blue line is a line of best fit (estimated using cubic regression splines), while the grey shaded zone represents its 95% confidence limits.

REFERENCES

1. 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.
2. Kristensen SD, Knuuti J, Saraste A, et al. 2014 ESC/ESA Guidelines on non-cardiac surgery: cardiovascular assessment and management: The Joint Task Force on non-cardiac surgery: cardiovascular assessment and management of the European Society of Cardiology (ESC) and the European Society of Anaesthesiology (ESA). *Eur Heart J* 2014; **35**: 2383–431.
3. Melon CC, Eshtiaghi P, Luksun WJ, Wijesundera DN. Validated questionnaire vs physicians' judgment to estimate preoperative exercise capacity. *JAMA Intern Med* 2014; **174**: 1507–8.
4. Reilly DF, McNeely MJ, Doerner D, et al. Self-reported exercise tolerance and the risk of serious perioperative complications. *Arch Intern Med* 1999; **159**: 2185–92.
5. Wiklund RA, Stein HD, Rosenbaum SH. Activities of daily living and cardiovascular complications following elective, noncardiac surgery. *Yale J Biol Med* 2001; **74**: 75–87.
6. Guazzi M, Adams V, Conraads V, et al. EACPR/AHA Scientific Statement. Clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. *Circulation* 2012; **126**: 2261–74.

7. Hlatky MA, Boineau RE, Higginbotham MB, et al. A brief self-administered questionnaire to determine functional capacity (the Duke Activity Status Index). *Am J Cardiol* 1989; **64**: 651–4.
8. Struthers A, Lang C. The potential to improve primary prevention in the future by using BNP/N-BNP as an indicator of silent ‘pancardiac’ target organ damage: BNP/N-BNP could become for the heart what microalbuminuria is for the kidney. *Eur Heart J* 2007; **28**: 1678–82.
9. Wijeyesundera DN, Pearse RM, Shulman MA, et al. Measurement of Exercise Tolerance before Surgery (METS) study: a protocol for an international multicentre prospective cohort study of cardiopulmonary exercise testing prior to major non-cardiac surgery. *BMJ Open* 2016; **6**: e010359.
10. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 2004; **240**: 205–13.
11. International Surgical Outcomes Study group. Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries. *Br J Anaesth* 2016; **117**: 601–9.
12. Thygesen K, Alpert JS, Jaffe AS, et al. Third universal definition of myocardial infarction. *Circulation* 2012; **126**: 2020–35.
13. Moran J, Wilson F, Guinan E, McCormick P, Hussey J, Moriarty J. Role of cardiopulmonary exercise testing as a risk-assessment method in patients undergoing intra-abdominal surgery: a systematic review. *Br J Anaesth* 2016; **116**: 177–91.

14. West MA, Asher R, Browning M, et al. Validation of preoperative cardiopulmonary exercise testing-derived variables to predict in-hospital morbidity after major colorectal surgery. *Br J Surg* 2016; **103**: 744–52.
15. 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.
16. Ford MK, Beattie WS, Wijeyesundera DN. Systematic review: prediction of perioperative cardiac complications and mortality by the Revised Cardiac Risk Index. *Ann Intern Med* 2010; **152**: 26–35.
17. Duceppe E, Parlow J, MacDonald P, et al. Canadian Cardiovascular Society guidelines on perioperative cardiac risk assessment and management for patients who undergo noncardiac surgery. *Can J Cardiol* 2017; **33**: 17–32.
18. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; **49**: 1373–9.
19. Pencina MJ, D’Agostino RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011; **30**: 11–21.
20. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; **148**: 839–43.
21. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology* 1990; **1**: 43–6.

22. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014; **35**: 1925–31.
23. Little RJA, Rubin DB. Statistical Analysis with Missing Data. 2nd Edition. Hoboken, NJ: *Wiley-Interscience*, 2002.
24. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009; **150**: 604–12.
25. Older P, Smith R, Courtney P, Hone R. Preoperative evaluation of cardiac failure and ischemia in elderly patients by cardiopulmonary exercise testing. *Chest* 1993; **104**: 701–4.
26. Young EL, Karthikesalingam A, Huddart S, et al. A systematic review of the role of cardiopulmonary exercise testing in vascular surgery. *Eur J Vasc Endovasc Surg* 2012; **44**: 64–71.
27. Carliner NH, Fisher ML, Plotnick GD, et al. Routine preoperative exercise testing in patients undergoing major noncardiac surgery. *Am J Cardiol* 1985; **56**: 51–8.
28. McPhail N, Calvin JE, Shariatmadar A, Barber GG, Scobie TK. The use of preoperative exercise testing to predict cardiac complications after arterial reconstruction. *J Vasc Surg* 1988; **7**: 60–8.
29. Forshaw MJ, Strauss DC, Davies AR, et al. Is cardiopulmonary exercise testing a useful test before esophagectomy? *Ann Thorac Surg* 2008; **85**: 294–9.
30. James S, Jhanji S, Smith A, O'Brien G, Fitzgibbon M, Pearse RM. Comparison of the prognostic accuracy of scoring systems, cardiopulmonary exercise testing, and plasma biomarkers: a single-centre observational pilot study. *Br J Anaesth* 2014; **112**: 491–7.

31. Whittle J, Nelson A, Otto JM, et al. Sympathetic autonomic dysfunction and impaired cardiovascular performance in higher risk surgical patients: implications for perioperative sympatholysis. *Open Heart* 2015; **2**: e000268.
32. Struthers R, Erasmus P, Holmes K, Warman P, Collingwood A, Sneyd JR. Assessing fitness for surgery: a comparison of questionnaire, incremental shuttle walk, and cardiopulmonary exercise testing in general surgical patients. *Br J Anaesth* 2008; **101**: 774–80.
33. Tang WH, Topol EJ, Fan Y, et al. Prognostic value of estimated functional capacity incremental to cardiac biomarkers in stable cardiac patients. *J Am Heart Assoc* 2014; **3**: e000960.
34. Grodin JL, Hammadah M, Fan Y, Hazen SL, Tang WH. Prognostic value of estimating functional capacity with the use of the duke activity status index in stable patients with chronic heart failure. *J Card Fail* 2015; **21**: 44–50.
35. Coutinho-Myrrha MA, Dias RC, Fernandes AA, et al. Duke Activity Status Index for cardiovascular diseases: validation of the Portuguese translation. *Arq Bras Cardiol* 2014; **102**: 383–90.
36. McIsaac DI, Bryson GL, van Walraven C. Association of frailty and 1-year postoperative mortality following major elective noncardiac surgery: a population-based cohort study. *JAMA Surg* 2016; **151**: 538–45.
37. Rodseth RN, Biccard BM, Le Manach Y, et al. The prognostic value of pre-operative and post-operative B-type natriuretic peptides in patients undergoing noncardiac surgery. B-type natriuretic peptide and N-terminal fragment of pro-B-type natriuretic peptide: a

- systematic review and individual patient data meta-analysis. *J Am Coll Cardiol* 2014; **63**: 170–80.
38. Botto F, Alonso-Coello P, Chan MT, et al. 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.
 39. Devereaux PJ, Xavier D, Pogue J, et al. Characteristics and short-term prognosis of perioperative myocardial infarction in patients undergoing noncardiac surgery: a cohort study. *Ann Intern Med* 2011; **154**: 523–8.
 40. Writing Committee for the VISION Study Investigators. Association of postoperative high-sensitivity troponin levels with myocardial injury and 30-day mortality among patients undergoing noncardiac surgery. *JAMA* 2017; **317**: 1642–51.
 41. Beattie WS, Wijesundera DN, Chan MTV, et al. Survival after isolated post-operative troponin elevation. *J Am Coll Cardiol* 2017; **70**: 907–8.
 42. Moonesinghe SR, Harris S, Mythen MG, et al. Survival after postoperative morbidity: a longitudinal observational cohort study. *Br J Anaesth* 2014; **113**: 977–84.
 43. Devereaux PJ, Mrkobra M, Sessler DI, et al. Aspirin in patients undergoing noncardiac surgery. *N Engl J Med* 2014; **370**: 1494–503.
 44. Myles PS, Smith JA, Forbes A, et al. Stopping vs. continuing aspirin before coronary artery surgery. *N Engl J Med* 2016; **374**: 728–37.
 45. 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.

46. Myles PS, Leslie K, Chan MT, et al. The safety of addition of nitrous oxide to general anaesthesia in at-risk patients having major non-cardiac surgery (ENIGMA-II): a randomised, single-blind trial. *Lancet* 2014; **384**: 1446–54.
47. Kopec M, Duma A, Helwani MA, et al. Improving prediction of postoperative myocardial infarction with high-sensitivity cardiac troponin T and NT-proBNP. *Anesth Analg* 2017; **124**: 398–405.
48. Ndumele CE, Matsushita K, Sang Y, et al. N-Terminal Pro-Brain Natriuretic Peptide and heart failure risk among individuals with and without obesity: The Atherosclerosis Risk in Communities (ARIC) Study. *Circulation* 2016; **133**: 631–8.
49. Gregg LP, Adams-Huet B, Li X, et al. Effect modification of chronic kidney disease on the association of circulating and imaging cardiac biomarkers with outcomes. *J Am Heart Assoc* 2017; **6**: e005235.
50. Tromp J, Khan MA, Klip IT, et al. Biomarker profiles in heart failure patients with preserved and reduced ejection fraction. *J Am Heart Assoc* 2017; **6**: e003989.
51. Fowler AJ, Ahmad T, Phull MK, Allard S, Gillies MA, Pearse RM. Meta-analysis of the association between preoperative anaemia and mortality after surgery. *Br J Surg* 2015; **102**: 1314–24.
52. Otto JM, O'Doherty AF, Hennis PJ, et al. Association between preoperative haemoglobin concentration and cardiopulmonary exercise variables: a multicentre study. *Perioper Med (Lond)* 2013; **2**: 18.

Table 1: Characteristics of study cohort

	Overall Cohort (N = 1401)	Missing Data (n)
Demographics		
Age – years, median (IQR)	65 (57 – 72)	
Female sex – no. (%)	548 (39%)	
Comorbidities		
Coronary artery disease	165 (12%)	
Heart failure	20 (1%)	
Cerebrovascular disease	59 (4%)	
Peripheral arterial disease	42 (3%)	
Diabetes mellitus	264 (19%)	
Hypertension	779 (56%)	
Current or recent smoker*	216 (15%)	
Obstructive lung disease†	181 (13%)	
Significant arthritis‡	289 (21%)	
Significant malignancy§	597 (43%)	
Preoperative renal function¶		49
eGFR ≥60 mL/min/1.73 m ² – no. (%)	1195 (88%)	
eGFR 30 – 59 mL/min/1.73 m ² – no. (%)	125 (9%)	
eGFR <30 mL/min/1.73 m ² or dialysis – no. (%)	31 (2%)	
Composite Risk Scales		
ASA-PS Classification		3
Class I	103 (7%)	
Class 2	818 (59%)	
Class 3	457 (33%)	
Class 4	20 (1%)	
Revised Cardiac Risk Index**		
Class I	624 (45%)	
Class 2	635 (45%)	
Class 3	115 (8%)	
Class 4	27 (2%)	
Preoperative Medications		
Beta-blocker	232 (17%)	
Dihydropyridine calcium channel blocker	248 (18%)	
Diltiazem or verapamil	27 (2%)	
ACE inhibitor or ARB	529 (38%)	

Furosemide	55 (4%)	
Aspirin	334 (24%)	
Other anti-platelet medication	35 (3%)	
Operative Characteristics		
Procedure type – no. (%)		
Vascular	26 (2%)	
Intra-thoracic	31 (2%)	
Intra-peritoneal or retro-peritoneal	464 (33%)	
Urologic or gynaecologic	417 (30%)	
Head-and-neck	93 (7%)	
Orthopaedic	336 (24%)	
Other	30 (2%)	
Laparoscopic or thoracoscopic assistance – no. (%)	499 (36%)	1
Anaesthesia type – no. (%)		
General anaesthesia alone	758 (54%)	
Regional anaesthesia alone	210 (15%)	
General plus regional anaesthesia	433 (31%)	
Intraoperative haemodynamic monitoring – no. (%)		
Arterial line	687 (49%)	7
Central venous line	212 (15%)	7
Cardiac output monitor	95 (7%)	8
Postoperative disposition		
Critical care unit or monitored bed unit	331 (24%)	1

Abbreviations: ACE, angiotensin converting enzyme; ARB, angiotensin receptor blocker; ASA-PS, American Society of Anesthesiologists Physical Status; eGFR, estimated glomerular filtration rate

* Current smoker or quit within previous one year

† Prior diagnosis of asthma, reactive airways disease, chronic obstructive lung disease, chronic bronchitis, or emphysema

‡ Previous or scheduled major joint replacement surgery

§ Indication for surgery was treatment of cancer

¶ Estimated glomerular filtration rate was calculated using the preoperative serum creatinine concentration and Chronic Kidney Disease Epidemiology Collaboration equation.²⁴

** Revised Cardiac Risk Index scores were calculated using the original definitions of diabetes mellitus (i.e., requirement for insulin therapy) and renal insufficiency (i.e., creatinine concentration >176 µmol/L) were employed.¹⁵ When determining Revised Cardiac Risk Index scores, any individual with missing preoperative creatinine concentration data was assumed as having a concentration ≤176 µmol/L

Table 2: Characteristics of CPET assessments

	Overall Cohort (N = 1401)
Interval to CPET to surgery – days, median (IQR)	9 (5 - 21)
Early termination of CPET – no. (%)	157 (11%)
Reasons for early termination of CPET – no. (%)	
Safety-based indication	23 (2%)
Fatigue	31 (2%)
Shortness of breath	11 (0.8%)
Unable to pedal	76 (5%)
Unable to tolerate mouthpiece or mask	12 (0.9%)
Technical problems with equipment	4 (0.3%)
Measurable VO ₂ peak – no. (%)	1356 (97%)
Measurable AT – no. (%)	1275 (91%)
Adverse events during CPET – no. (%)	
Any adverse event	110 (8%)
Ischaemic ECG changes	25 (2%)
Chest pain	2 (0.1%)
Significant new arrhythmias	14 (1%)
Significant hypertension	43 (3%)
Significant drop in blood pressure	28 (2%)
Syncope	3 (0.2%)
Significant drop in oxygen saturation	10 (0.7%)

Abbreviations: AT, anaerobic threshold; CPET, cardiopulmonary exercise testing; ECG, electrocardiogram; VO₂ peak, peak oxygen consumption

Table 3: Predictive performance of different measures of preoperative functional capacity

	Adjusted Odds Ratio (95% CI)	AUC*	Net Reclassification Improvement Index [†]		
			Events	Non-Events	Overall
30-Day Death or Myocardial Infarction					
Baseline model [‡]		0.59			
+VO ₂ peak	aOR 0.90 (0.71-1.16; p=0.45) [§]	0.62	0.04	-0.04	0.01 (p=0.98)
+AT	aOR 0.96 (0.66-1.41; p=0.84) [§]	0.59	-0.24	-0.12	-0.36 (p=0.10)
+DASI	aOR 0.91 (0.83-0.99; p=0.03) [§]	0.67	0.07	0.21	0.28 (p=0.14)
+NT pro-BNP	aOR 1.88 (0.89-3.96; p=0.09) [¶]	0.65	0.11	0.14	0.25 (p=0.19)
30-Day Death or Myocardial Injury					
Baseline model ^{**}		0.70			
+VO ₂ peak	aOR 1.03 (0.92-1.14; p=0.62) [§]	0.70	-0.07	0.16	0.09 (p=0.26)
+AT	aOR 1.12 (0.96-1.31; p=0.16) [§]	0.71	0.17	-0.08	0.09 (p=0.27)
+DASI	aOR 0.96 (0.92-0.99; p=0.05) [§]	0.71	0.05	0.19	0.23 (p=0.004)
+NT pro-BNP	aOR 1.78 (1.21-2.62; p=0.003) [¶]	0.71	0.07	0.13	0.20 (p=0.02)
In-Hospital Moderate-or-Severe Complications					
Baseline model ^{††}		0.72			
+VO ₂ peak	aOR 0.86 (0.78-0.97; p=0.007) [§]	0.74	0.21	-0.004	0.21 (p=0.008)
+AT	aOR 0.87 (0.74-1.02; p=0.08) [§]	0.69	0.17	-0.08	0.09 (p=0.27)
+DASI	aOR 0.97 (0.93-1.01; p=0.16) [§]	0.72	-0.17	0.15	-0.02 (p=0.80)
+NT pro-BNP	aOR 1.10 (0.77-1.57; p=0.61) [¶]	0.72	0.03	0.08	0.12 (p=0.12)
1-Year Death					

Baseline model [‡]		0·65			
+VO ₂ peak	aOR 0·94 (0·77-1·15; p=0·56) [§]	0·66	0·20	-0·06	0·14 (p=0·39)
+AT	aOR 1·03 (0·76-1·40; p=0·56) [§]	0·64	0·13	0·15	0·28 (p=0·12)
+DASI	aOR 0·94 (0·87-1·02; p=0·13) [§]	0·69	0·00	0·16	0·16 (p=0·34)
+NT pro-BNP	aOR 2·91 (1·54-5·49; p=0·001) [¶]	0·72	0·17	0·23	0·39 (p=0·02)

Abbreviations: AUC, area under the curve; AT, anaerobic threshold; aOR, adjusted odds ratio; CI, confidence interval; VO₂ peak, peak oxygen consumption

* Area under the curve of the receiver operating characteristic curve for the relevant logistic regression model.

† The net classification index is the weighed net proportion of individuals whose predicted probability of the outcome of interest improved with inclusion of the specific additional covariate in the regression model (e.g., VO₂ peak). Improved predicted probability implies a higher predicted probability in individuals who had the outcome event of interest, and a lower predicted probability in individuals who did not. Negative statistic values indicate net worsening of predicted probabilities (e.g., higher predicted probability in individuals who did not have the outcome of interest). This statistic has an associated p-value.

‡ The covariate in this baseline model was the Revised Cardiac Risk Index score.

§ Adjusted odds ratios were expressed with respect to 1 metabolic equivalent increase in VO₂ peak (per 3·5 mL/kg/min), AT (per 3·5 mL/kg/min), or DASI scores (per 3·5 points).

¶ Adjusted odds ratios are expressed with respect to 1 log₁₀ increase in NT pro-BNP concentrations

** The covariates in this baseline model were age, sex and Revised Cardiac Risk Index score.

†† The covariates in this baseline model were age, sex, and high-risk surgery (intra-peritoneal, intra-thoracic, or supra-inguinal vascular procedures).



