Sex-related differences in functional capacity and its implications in risk stratification before major non-cardiac surgery: a *post hoc* analysis of the international METS study

Jonas Alfitian,[a,l,*] Bernhard Riedel,[b,c,d,l] Hilmy Ismail,[b,c] Kwok M. Ho,[e] Sophia Xie,[f] Philipp Zimmer,[g] Tobias Kammerer,[a] Duminda N. Wijeysundera,[h,i] Brian H. Cuthbertson,[h,j] and Robert Schier,[a,k] on behalf of the METS Study Investigators

[a]University of Cologne, Faculty of Medicine and University Hospital Cologne, Department for Anesthesiology and Intensive Care Medicine, Germany
[b]Department of Anaesthesia, Perioperative Medicine and Pain Medicine, Peter MacCallum Cancer Centre, Australia
[c]The Department of Critical Care, University of Melbourne, Melbourne, Australia
[d]The Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Australia
[e]University of Western Australia and Murdoch University, Australia
[f]Peter MacCallum Cancer Centre, Centre for Biostatistics and Clinical Trials, Australia
[g]Division of Performance and Health, Institute for Sport and Sport Science, TU Dortmund University, Germany
[h]Department of Anesthesiology and Pain Medicine, University of Toronto, Toronto, ON, Canada
[i]Department of Anesthesia, St. Michael's Hospital, Toronto, ON, Canada
[j]Department of Critical Care Medicine, Sunnybrook Health Sciences Centre, Toronto, ON, Canada
[k]Department of Anesthesiology, Emergency and Intensive Care Medicine, University of Marburg, Campus Fulda, Germany

eClinicalMedicine
2023;64: 102223
Published Online xxx
https://doi.org/10.1016/j.eclinm.2023.102223

Summary

Background Poor functional capacity has been identified as an important modifiable risk factor for postoperative complications. Cardiopulmonary exercise testing (CPET) provides objective parameters of functional capacity (e.g., oxygen consumption at peak exercise, peak VO_2), with significant prognostication for postoperative complications. However, sex-specific thresholds for functional capacity to predict surgical risk are yet to be established. Therefore, we performed a *post hoc* analysis of the international, multicentre, prospective observational METS (Measurement of Exercise Tolerance before Surgery) study to evaluate if sex-specific thresholds of peak VO_2 improve risk prediction of postoperative complications.

Methods We undertook a *post hoc* analysis (HREC/71824/PMCC) of the METS study, which was performed between March 2013 and March 2016. We investigated whether sex-specific differences exist for CPET-derived parameters and associated thresholds for predicting postoperative complications in this large cohort of patients that had major non-cardiac surgery (n = 1266). Logistic regression models were analyzed for the association of low peak VO_2 with moderate-to-severe in-hospital postoperative complications. Optimal sex-specific peak VO_2 thresholds were obtained by maximizing the Youden index of receiver operating characteristic (ROC) curves. Finally, multivariable logistic regression models tested the resulting sex-specific thresholds against the established non-sex-specific peak VO_2 threshold (14 mL kg^{-1} min^{-1}) adjusted for clinically relevant features such as comorbidities and surgical complexity. Models were evaluated by bootstrapping optimism-corrected area under the ROC curve and the net reclassification improvement index (NRI).

Findings Female patients (n = 480) had a lower mean (SD) peak VO_2 than males (16.7 (4.9) mL kg^{-1} min^{-1} versus 21.2 (6.5) mL kg^{-1} min^{-1}, $p < 0.001$) and a lower postoperative complication rate (10.4% versus 15.3%; p = 0.018) than males (n = 786). The optimal peak VO_2 threshold for predicting postoperative complications was 12.4 mL kg^{-1} min^{-1} for females and 22.3 mL kg^{-1} min^{-1} for males, respectively. In the multivariable regression model, low non-sex-specific peak VO_2 did not independently predict postoperative complications. In contrast, low sex-specific peak VO_2 was an independent predictor of postoperative complications (OR 2.29; 95% CI: 1.60, 3.30; $p < 0.001$). The optimism-corrected AUC-ROC of the sex-specific model was higher compared with the non-sex-specific model (0.73 versus 0.7; DeLong's test: p = 0.021). The sex-specific model classified 39% of the patients more correctly than the baseline model (NRI = 0.39; 95% CI: 0.24, 0.55). In contrast, the non-sex-specific

*Corresponding author.
E-mail addresses: jonas.alfitian@uk-koeln.de (J. Alfitian), bernhard.riedel@petermac.org (B. Riedel).
[l]These authors contributed equally to this work.

model only classified 9% of the patients more correctly when compared against the baseline model (NRI = 0.09; 95% CI: −0.04, 0.22).

Interpretation Our data report sex-specific differences in preoperative CPET-derived functional capacity parameters. Sex-specific peak VO_2 thresholds identify patients at increased risk for postoperative complications with a higher discriminatory ability than a sex-unspecific threshold. As such, sex-specific threshold values should be considered in preoperative CPET to potentially improve risk stratification and to guide surgical decision-making, including eligibility for surgery, preoperative optimization strategies (prehabilitation) or seeking non-surgical options.

Funding There was no funding for the present study. The original METS study was funded by 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, and Monash University.

Keywords: Cardiopulmonary exercise testing; Peak oxygen uptake; Gender medicine; Non-cardiac surgery; Postoperative complications

Research in context

Evidence before this study

The objective assessment of a patient's functional capacity with cardiopulmonary exercise testing (CPET) before major non-cardiac surgery is widely used for surgical risk assessment. Importantly, poor functional capacity is a modifiable risk factor, with increasing evidence that exercise-based prehabilitation before surgery significantly reduces postoperative complications and hospital length of stay. A universal threshold of increased risk is typically set at four metabolic equivalents (4 METS, peak VO_2 14 mL kg^{-1} min^{-1}) for cardiovascular risk assessment, as per the current clinical guidelines, such as the American College of Cardiology and the American Heart Association (ACC/AHA) guidelines. However, sex-specific differences in exercise capacity have not been investigated in the perioperative setting for predicting postoperative complications after major non-cardiac surgery. We postulated that sex-specific differences exist in patients' functional capacity and that a single non-sex-specific threshold (e.g., 4 METS) will likely overestimate the risk in females and underestimate the risk in males, potentially leading to disparity in healthcare.

Added value of this study

This *post hoc* analysis of the METS study contributes to precision medicine by recognising that sex-related differences in functional capacity exist in preoperative risk prediction when using CPET-derived parameters such as peak VO_2. This improves risk stratification and, thereby, the guidance of surgical decision-making, including delaying surgery for preoperative optimization strategies such as exercise-based prehabilitation.

Implications of all the available evidence

Numerous studies have shown that low functional capacity is associated with an increased risk of postoperative complications in non-cardiac surgery and that preoperative interventions, such as exercise-based prehabilitation, modify this risk. We provide evidence that sex-adjusted interpretation of CPET-derived functional capacity parameters such as peak VO_2 may improve risk prediction and, thereby, the guidance for surgical decision-making. Specifically, we demonstrate improved risk prediction for high-grade postoperative complications when considering sex-specific peak VO_2 thresholds compared to the established standard of care of a non-sex-specific CPET threshold set at 4 METS (14 mL kg^{-1} min^{-1}) that is widely accepted in perioperative risk prediction guidelines.

Introduction

Mortality within 30 days after surgery is the third leading cause of death globally, and postoperative complications primarily drive it.[1] Poor functional capacity has been identified as an important modifiable risk factor for postoperative complications.[2] This necessitates improved preoperative surgical risk stratification and optimization strategies. Preoperative cardiopulmonary exercise testing (CPET) using cycle ergometry with spirometry provides objective parameters of functional capacity (e.g., oxygen consumption at anaerobic threshold [AT] or at peak exercise [peak VO_2]), which allows for prognostication for postoperative complications and for guiding preoperative exercise-based optimization strategies (prehabilitation).[3] Older et al.[4,5] were the first to correlate CPET testing with mortality in patients undergoing major abdominal surgery in the early 1990s. Subsequently, numerous studies confirmed that CPET-derived functional capacity parameters independently predict postoperative morbidity and mortality in patients scheduled for major abdominal surgery, including colorectal, upper gastrointestinal, hepatobiliary, urological, and abdominal aortic aneurysm surgery.[6–18] Importantly, these studies failed to consider the differences in functional capacity attributed to biological sex.

Sex-specific differences are the subject of contemporary clinical practice and medical research, as biological sex is a genetic modifier of disease pathophysiology and response to medical treatment. Sex differences with exercise testing have been reported, with sex-related differences in the progression of metabolic responses during incremental exercise.[19] Similarly, one study reported that functional capacity in men and chronotropic competence in women are the primary prognosticators for longevity, respectively.[18] Thomas et al.[20] reported that CPET-derived AT and peak VO_2 differed significantly between males and females, however, they did not investigate its implications for postoperative outcome. As such, little is known about sex-related differences in interpreting CPET-derived parameters and their risk stratification threshold for postoperative complications after major non-cardiac surgery. This necessitates deeper investigation to promote sex equity in health care, as there may be different thresholds for increased surgical risk that guide the need for surgical deferment for prehabilitation or the consideration of alternative therapy to surgery.

In 2018, the prospective, international, multicentre METS (Measurement of Exercise Tolerance before Surgery) study was published in *The Lancet*.[2] This is the largest (n = 1401 patients) observational study assessing the association of functional capacity with postoperative risk—with death or myocardial infarction within 30 days of surgery investigated as the primary outcome. Notably, the physician estimate of patients' functional capacity was very poor. The patient reported Duke Activity Status Index (DASI) score was the only parameter associated with the primary outcome of death or myocardial infarction by postoperative day 30. Significantly, CPET-derived parameters, specifically peak VO_2, did associate with increased risk for moderate or severe in-hospital postoperative complications and hence reinforces the importance of objective functional capacity assessment before major non-cardiac surgery for surgical risk assessment. The importance of this is further underpinned by the fact that poor functional capacity is a modifiable risk factor, with an increasing number of studies of exercise-based prehabilitation reporting a significant reduction, by as much as half, of postoperative complications[21] and a reduction in hospital length of stay, by as much as three days.[22] Consequently, cardiopulmonary exercise testing for preoperative surgical risk assessment is well established in the United Kingdom, with more than 50% of hospitals providing preoperative CPET facilities and prehabilitation services.[3]

Thresholds for CPET-derived functional parameters allow for risk stratification that guides clinical decision-making. A widely accepted threshold of increased surgical risk is typically set at four metabolic equivalents (4 METs, peak VO_2 14 mL kg^{-1} min^{-1}) and primarily for cardiac risk prediction.[23] This threshold reflects low cardiac performance within guidelines set for patients with chronic heart failure.[24,25] Given that sex-specific differences in exercise capacity have not been investigated in the non-cardiac surgery setting, we postulated that sex-specific differences exist in patients' functional capacity. As the METS study has already demonstrated that peak VO_2 is related to the risk of moderate-to-severe complications, the present work focuses on the role of sex regarding this relationship. We set out to investigate whether sex-specific thresholds for peak VO_2 are more suitable than the non-sex-specific threshold of 14 mL kg^{-1} min^{-1} to predict moderate-to-severe complications after major non-cardiac surgery. Furthermore, considering the common use of questionnaire-based assessment of METs, our secondary objective set out to examine sex-specific differences among the relationship between Duke Activity Status Index (DASI) and both AT and peak VO_2.

Methods

Study design and participants

We conducted a *post hoc* analysis (HREC/71824/PMCC) of the METS study,[2] which was performed between March 2013 and March 2016. The METS study compared preoperative subjective assessment by physicians with alternative markers of fitness (CPET, scores of the DASI questionnaire, and serum N-terminal pro-B-type natriuretic peptide [NT pro-BNP] concentrations) for predicting the primary endpoint of myocardial infarction, death within 30 days after surgery or a secondary endpoint of complications after major elective non-cardiac surgery. The METS study was conducted as a prospective multicenter study, with patients enrolled from 25 hospitals in Australia, Canada, New Zealand, and the United Kingdom, between 2013 and 2016. Eligible patients were scheduled for major non-cardiac surgery with at least one overnight stay, were aged 40 years or older, with at least one risk factor for a cardiac complication (diabetes, coronary artery disease, cerebrovascular disease, renal dysfunction, peripheral arterial disease, hypertension, smoker, aged over 70 years). Exclusion criteria were inadequate time to complete CPET before surgery (defined as <24 h), planned use of CPET for preoperative risk stratification independent of METS study protocol, planned surgery exclusively performed by an endovascular approach, presence of an automated implantable cardioverter-defibrillator, known or suspected pregnancy, previous enrollment in the METS Study, absolute contraindications to CPET (according to American Thoracic Society and American College of Chest Physicians guidelines[26]), conditions expected to preclude CPET (e.g., lower limb amputation, severe claudication), systolic blood pressure ≥180 mmHg and diastolic blood pressure ≥100 mmHg at the time of study recruitment.

Study objectives

Within our *post hoc* analysis, our main objectives were primarily whether functional capacity, specifically peak VO_2, resulted in sex-specific differences with regards to preoperative risk stratification. Therefore, classification models were built with subsequent determination of optimal sex-specific thresholds for predicting postoperative complications. Sex-specific thresholds were then compared with an established sex-unspecific threshold in terms of prediction accuracy. Our secondary objective was to investigate how biological sex affects the correlation of DASI and both the actual CPET-derived AT and peak VO_2 value.

Assessment of functional capacity

Functional capacity was objectively assessed by measuring AT and peak VO_2 using CPET subjectively using the DASI questionnaire. Both procedures were performed between the time of recruitment and one day before scheduled surgery. Peak VO_2 and AT were measured by medical-supervised CPET that used a symptom-limited computer-controlled ramp protocol with an electromagnetically braked cycle ergometer and according to international guidelines.[26] A 3-min phase of seated rest was followed by a 3-min phase of unloaded cycling (0 W). In the subsequent loading phase, work load was gradually increased by approximately 10–30 W per minute at 60–70 revolutions per minute (rpm) depending on patient performance until pedaling at 60 rpm became unsustainable thus reaching tolerance limit. During the entire procedure, besides heart rate, blood pressure and ECG were recorded. Breath-by-breath analysis allowed calculation of minute ventilation, oxygen uptake and carbon dioxide production. The evaluation of CPET was blinded and peak VO_2 was obtained at the time of the tolerance limit. AT was determined using the modified V-slope method.[27] Patients with missing data on peak VO_2 or AT were excluded from our analysis.

For our secondary objective, self-reported functional capacity was measured via the DASI using a 12-item questionnaire for activities of daily living that has been correlated with peak VO_2.[28] Study participants were asked to complete the DASI questionnaire at the time of study recruitment.

Main outcome measures

All study participants were visited daily during their hospital stay. For our study, the primary outcome of interest was the occurrence of moderate-to-severe in-hospital postoperative complications, which was reported in the METS study and other studies to be significantly associated with impaired functional capacity objectively measured as peak VO_2 during CPET.[2,29,30] Complications were graded using an adaption of Clavien-Dindo scoring system.[31,32] Mild complications were defined as those resulting in temporary harm that would not usually require specific clinical treatment. Moderate complications were defined as more serious complications requiring clinical treatment but did not result in permanent harm or functional limitation. Severe complications were defined as those that required clinical treatment and resulted in significant prolongation of hospital stay and/or permanent functional limitation. Types of complications were categorized into myocardial infarction (according to Third Universal Definition of Myocardial Infarction[33]), myocardial injury (defined as postoperative troponin concentrations exceeding 99th percentile of the normal reference population), non-fatal cardiac arrest, congestive heart failure, stroke or transitory ischemic attack, respiratory failure, pneumonia, surgical site infection, deep venous thrombosis, pulmonary embolism, and re-operation according to their definition outlined in the original publication of the METS study.[2] Patients with missing data on postoperative complications were excluded from our analysis.

Building of regression models

The association of peak VO_2 with moderate-to-severe complications as a binary outcome was examined using logistic regression models. In a first step, a baseline model was built to adjust for clinically relevant variables for adverse postoperative events. Here, we considered patient age, high-risk surgery (open surgical intraperitoneal, retroperitoneal, intrathoracic, or supra-inguinal vascular procedure), body mass index (BMI), chronic obstructive pulmonary disease (COPD), hypertension, coronary artery disease (CAD), cerebrovascular disease (CVD), and heart failure. Orthopedic surgery was considered an additional explanatory variable in the baseline model, as these patients may have limitations in their capacity to perform CPET due to joint pain and mobility issues, rather than solely reflecting differences in cardiovascular functional capacity. The baseline model was then extended with peak VO_2 as a continuous variable to determine its adjusted odds ratio for the outcome of postoperative complications and to investigate whether an unbiased relationship existed between functional capacity and postoperative complications before performing dichotomization of peak VO_2. This continuous peak VO_2 model was then further applied separately for females and males to test for an association of continuous peak VO_2 with the outcome of postoperative complications. To determine the optimal peak VO_2 thresholds, receiver operating characteristic (ROC) curves were generated separately for females and males. Subsequently, the Youden index[34] was calculated for each possible peak VO_2 threshold. The optimal peak VO_2 threshold for each biological sex was obtained by finding the maximum Youden index. Individuals were then classified into low versus normal peak VO_2 according to (1) the computed sex-specific threshold and (2) the widely accepted threshold of 14 mL kg^{-1} min^{-1}. To compare sex-specific peak VO_2 thresholds versus the sex-unspecific peak VO_2 threshold for prediction of postoperative complications, two distinct logistic regression models adjusted to the baseline model with low peak VO_2 as a categorical explanatory variable either by the sex-specific or by the sex-unspecific peak VO_2 threshold were built and applied on the full study population.

For the secondary objective, we investigated how biological sex affects the linear relationship between DASI score and both actual measured peak VO_2 and AT. Therefore, multivariable linear regression models were built utilizing peak VO_2 or AT as response variables and an interaction term between DASI score and biological sex as explanatory variables. This determined how biological sex influenced the steepness of the linear relationship of DASI score and actual peak VO_2 or AT, respectively. Additionally, we compared the DASI-predicted peak VO_2 (formula: peak $VO_2 = 0.43 \times \text{DASI} + 9.6$)[28] with the actual CPET-derived peak VO_2.

Statistical analysis

All statistical tests and regression model development was carried out in R version 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria).[35] Categorical variables were compared using the Pearson's chi-squared test. The two-sample t-test was applied for independence testing of two sample means. Logistic regression models were evaluated by ROC curve analysis with subsequent calculation of the area under the curve (AUC-ROC). Internal validation of AUC-ROC was achieved by the optimism-corrected bootstrapping method. Optimism-corrected AUC-ROC were compared by DeLong's test. Moreover, discriminatory power of classification models was assessed by computing the continuous Net Reclassification Improvement Index (NRI).[36] As for the AUC-ROC, the NRI was also validated internally via bootstrapping. Model calibration was tested using the Hosmer–Lemeshow test and moreover assessed by calibration plots including bias-corrected calibration curves that were computed by bootstrapping. All bootstrapping approaches were performed with 2000 replicates each. All statistical tests were two-tailed, and a p-value below 0.05 was considered significant.

Ethics statement

Ethical approval was obtained on January 14, 2021 (Peter MacCallum Cancer Centre Human Research Ethics Committee, Melbourne, Australia: #20/246L, January 14, 2021). All study participants provided informed written consent.

Role of the funding sources

JA, BR, HI, KH, SX, DNW, BHC and RS had access to the dataset. Funders had no role in data collection, data analysis, interpretation, trial design, patient recruitment or any aspect pertinent to the study. JA, BR and RS accept responsibility to submit for publication.

Results

Study population

The original population of the METS study comprised 1401 eligible patients who met the inclusion criteria according to the study protocol and underwent both CPET and surgery. In our *post hoc* analysis, we excluded patients with incomplete CPET data or missing information on postoperative complications within the first 30 days after surgery. Consequently, 1266 (90.4% of the original study) patients were included, with the study cohort having a median age of 65 years and comprising 37.9% female patients. Relevant preoperative risk scores, stratified by sex, including the American Society of Anesthesiologists (ASA) Physical Status and Revised Cardiac Risk Index (RCRI), data on comorbid disease, and surgery type are reported in Table 1. High-risk surgery was performed in 475 (37.5%) patients.

Postoperative outcome

During the hospital stay and within the first 30 days after surgery, one (0.2%) female and three (0.4%) male patients died. The severity and types of postoperative complications are summarised in Table 2. The primary endpoint of moderate-to-severe complications during hospital stay occurred less frequent in females than males (10.4% [n = 50] versus 15.3% [n = 120], respectively; p = 0.014).

Sex-specific differences of peak VO_2

We observed sex-specific differences in the parameters, with both AT and peak VO_2 being significantly lower in female patients (Table 3). There were no differences regarding ventilatory equivalents for oxygen (V_E/VO_2 ratio) and carbon dioxide (V_E/VCO_2 ratio). Poor functional capacity was regarded as functional capacity less than four METs (peak VO_2 14 mL kg^{-1} min^{-1}). Of the female patients, 130 (27.1%) had a peak VO_2 below this threshold, which was significantly more frequent than in males (63 patients; 8%) being below this traditional threshold ($p < 0.001$). Distribution of peak VO_2 stratified by biological sex and severity of postoperative complications is demonstrated in Fig. 1B.

We built three nested logistic regression models. The baseline model was used for adjusting for clinically relevant features. In this baseline model, only surgical properties were independent predictors of moderate-to-severe complications with orthopedic surgery patients being at lower risk in contrast to high-risk surgery patients, which were more likely to experience the

	Female, N = 480 (38%)	Male, N = 786 (62%)	Missing data
Demographics			
Age (years)	62.3 (11.0)	64.9 (9.8)	–
Body weight (kg)	77.3 (19.3)	85.7 (17.9)	–
BMI (kg/m^2)	29.5 (6.9)	28.4 (6.0)	–
Composite Risk Scores			
ASA status[a]			2 (0.2%)
Class 1	37 (7.7%)	62 (7.9%)	
Class 2	294 (61%)	453 (58%)	
Class 3	140 (29%)	259 (33%)	
Class 4	9 (1.9%)	10 (1.3%)	
Revised Cardiac Risk Index (RCRI)[b]			44 (3.5%)
Class 0	231 (49%)	301 (40%)	
Class 1	213 (45%)	359 (48%)	
Class 2	21 (4.5%)	69 (9.2%)	
Class 3	3 (0.6%)	21 (2.8%)	
Class 4	1 (0.2%)	3 (0.4%)	
Duke activity status index (DASI)	38.4 (14.4)	45.0 (13.7)	55 (4.3%)
Surgical properties			
Surgery type			–
Head-and-neck surgery	39 (8.1%)	46 (5.9%)	
Intra-peritoneal or retro-peritoneal	156 (32%)	269 (34%)	
Intra-thoracic	15 (3.1%)	16 (2.0%)	
Orthopedic	138 (29%)	140 (18%)	
Other	18 (3.8%)	14 (1.8%)	
Urologic or gynecologic surgery	114 (24%)	282 (36%)	
Vascular	0 (0%)	19 (2.4%)	
History of comorbidities			
Diabetes mellitus	69 (14%)	158 (20%)	–
Coronary artery disease	24 (5.0%)	131 (17%)	–
Cerebrovascular disease	17 (3.5%)	33 (4.2%)	–
Heart failure	6 (1.3%)	12 (1.5%)	–
Peripheral artery disease	9 (1.9%)	25 (3.2%)	–
Arterial hypertension	271 (56%)	413 (53%)	–

[a]ASA 1: normal healthy patient, ASA 2: mild systemic disease, ASA 3: severe systemic disease, ASA 4: severe systemic disease that is a constant threat to life. [b]RCRI is calculated by existence of the following risk factors, assigning one point for each: High-risk surgery (intraperitoneal, intrathoracic, supra-inguinal vascular), history of ischemic heart disease, history of congestive heart failure, history of cerebrovascular disease, preoperative treatment with insulin, preoperative serum creatinine >2 mg/dL or 176.8 μmol/L.

Table 1: **Characteristics of the study population.**

outcome of postoperative complications (Table 4). After the extending the baseline model by peak VO_2 as a continuous variable, this parameter was also an independent predictor of the likelihood of experiencing moderate-to-severe complications (OR = 0.95 per unit increase, 95% CI: 0.92, 0.98; p = 0.0011). The optimal peak VO_2 cut point was evaluated after applying this continuous peak VO_2 model to males and females separately. Maximum Youden index corresponded to a threshold of 12.4 mL kg^{-1} min^{-1} for females and 22.3 mL kg^{-1} min^{-1} for males. The ratio of males and females with a critically reduced peak VO_2 was reversed after applying these sex-specific threshold values, in contrast to the non-sex-specific threshold value of 14 mL kg^{-1} min^{-1}. Significantly, 66.7% of male patients showed a critically decreased sex-specific peak VO_2 compared to 16.7% of female patients ($p < 0.001$). To compare the predictive performance of low peak VO_2 based on the non-sex-specific versus sex-specific thresholds, we built two logistic regression models, each as an extension of the baseline model. The sex-unspecific model included low peak VO_2 < 14 mL kg^{-1} min^{-1} as a predictor, and in the sex-specific model, the corresponding computed threshold value for each sex was used to define low peak VO_2. Notably, a reduced peak VO_2 was an independent predictor of moderate-to-severe complications only when sex-specific thresholds were applied and not when determined by the non-sex-specific threshold value (Table 4). The optimism-corrected AUC-ROC of the sex-specific model was 0.73

	Female, N = 480	Male, N = 786	p-value[a]	Missing data
Fatal outcome				
Death within 30 days	1 (0.2%)	3 (0.4%)	>0.9	1 (0.1%)
Death during hospital stay	1 (0.2%)	3 (0.4%)	>0.9	–
Postoperative complications				
Severity of postoperative complications[b]			0.2	–
None	402 (84%)	617 (78%)		
Mild	28 (5.8%)	49 (6.2%)		
Moderate	38 (7.9%)	89 (11%)		
Severe	11 (2.3%)	28 (3.6%)		
Fatal	1 (0.2%)	3 (0.4%)		
None or mild	430 (90%)	666 (85%)	0.014	–
Moderate-to-severe	50 (10%)	120 (15%)	0.014	–
Type of complication				
Myocardial infarction	6 (1.3%)	12 (1.5%)	0.7	–
Myocardial injury	56 (12%)	92 (12%)	>0.9	–
Non-fatal cardiac arrest	0 (0%)	2 (0.3%)	0.5	–
Congestive heart failure	1 (0.2%)	0 (0%)	0.4	–
Stroke or TIA	0 (0%)	1 (0.1%)	>0.9	–
Respiratory failure	4 (0.8%)	12 (1.5%)	0.3	–
Pneumonia	5 (1.0%)	15 (1.9%)	0.2	–
Surgical site infection	6 (1.3%)	22 (2.8%)	0.069	–
Deep venous thrombosis	0 (0%)	1 (0.1%)	>0.9	–
Pulmonary embolism	1 (0.2%)	2 (0.3%)	>0.9	–
Re-operation	8 (1.7%)	22 (2.8%)	0.2	–
New requirement for dialysis	0 (0%)	3 (0.4%)	0.3	–
Unexpected ICU admission	15 (3.1%)	32 (4.1%)	0.4	–

[a]p-values were calculated using Pearson's Chi-squared test. [b]Severity of complications was assessed by an adaption of Clavien-Dindo classification.

Table 2: **Postoperative outcome.**

and thus higher than that of the non-sex-specific model with 0.7, with higher discriminatory ability of the sex-specific model as verified by DeLong's test (p = 0.021). The non-sex-specific model improved the net reclassification of patients for the primary endpoint compared to the baseline model with 9% of patients classified more correctly, while patients with the outcome were even classified worse as revealed by a negative event NRI (event NRI = –0.6, non-event NRI = 0.69, overall NRI = 0.09, 95% CI for overall NRI –0.04, 0.22).

	Female, N = 480	Male, N = 786	p-value[a]	Missing data
Peak VO_2 (mL kg^{-1} min^{-1})			<0.001	–
Mean (SD)	16.7 (4.9)	21.2 (6.5)		
Median (IQR)	16.0 (5.7)	20.0 (7.4)		
VO_2 at AT (mL kg^{-1} min^{-1})			<0.001	–
Mean (SD)	11.4 (3.5)	13.4 (4.3)		
Median (IQR)	11.0 (4.0)	12.8 (4.1)		
V_E/VCO_2 at AT			0.022	2 (0.2%)
Mean (SD)	32.4 (6.1)	31.5 (5.3)		
Median (IQR)	31.2 (8.0)	31.0 (6.0)		
V_E/VO_2 at AT			0.044	2 (0.2%)
Mean (SD)	29.4 (6.6)	28.7 (5.6)		
Median (IQR)	29.0 (8.0)	28.0 (6.0)		

AT, anaerobic threshold; IQR, Interquartile Range; SD, Standard Deviation V_E/VCO_2, ventilatory equivalent for CO_2; V_E/VO_2, ventilatory equivalent for O_2. [a]p-values were calculated using Welch's t-test.

Table 3: **Cardiopulmonary exercise testing-derived parameters of functional capacity and reserve.**

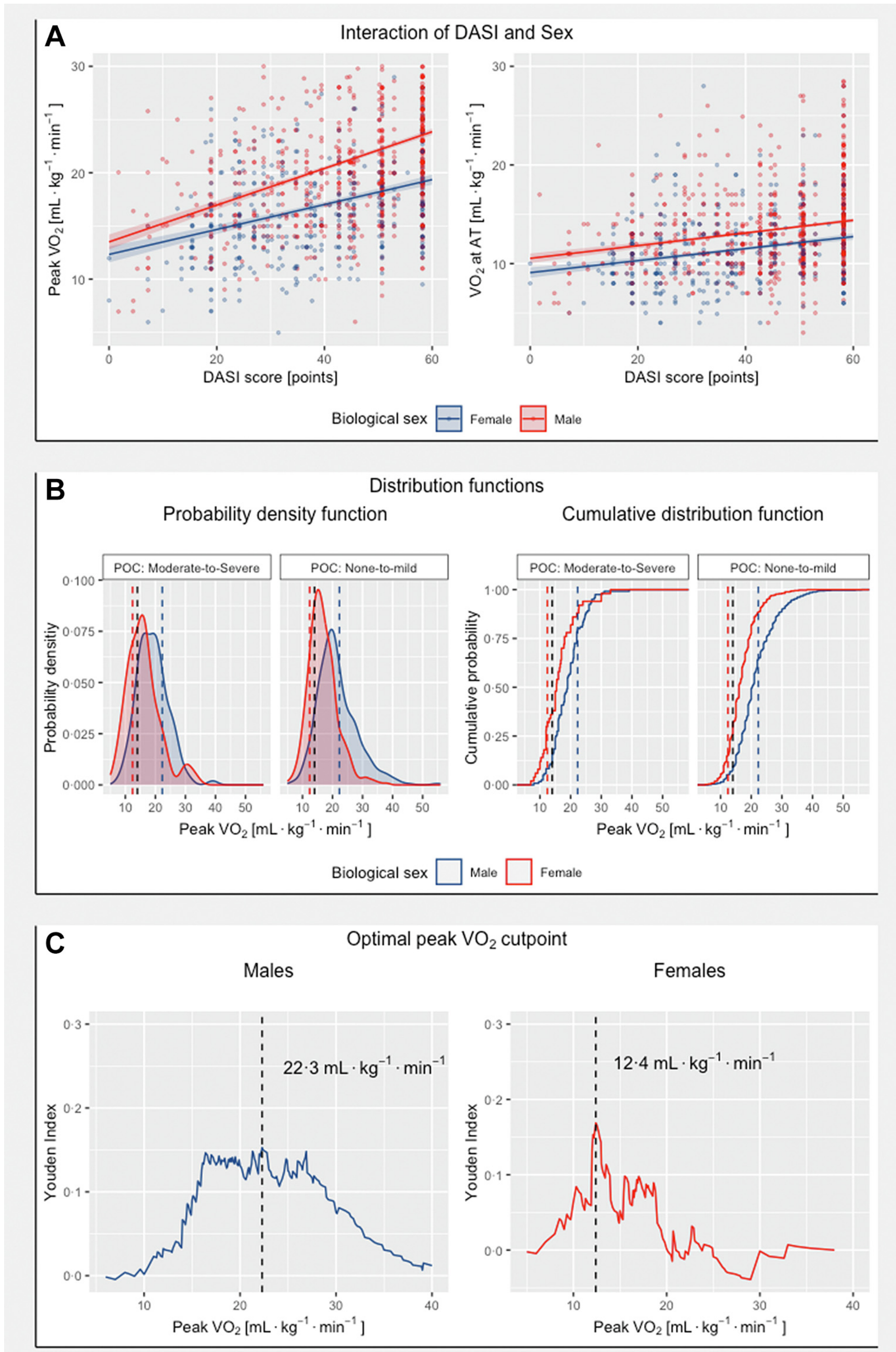

Fig. 1: Sex-related differences in subjective and objective functional capacity. A: The proportionality between subjective functional capacity reported via the DASI questionnaire and the objectifiable counterpart mapped as peak VO_2 (left panel) differs significantly between males and females, but not for oxygen consumption at AT (right panel). The graphs result from linear regression analysis with an interaction term between DASI and biological sex (details in Table A-1). Accordingly, peak VO_2 increases steadily and more steeply with increasing DASI in males

Predictor	Baseline model			Non-sex-specific peak VO_2 model			Sex-specific peak VO_2 model		
	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Age (years)	1.01	1.00, 1.03	0.157	1.01	0.99, 1.03	0.200	1.00	0.99, 1.02	0.634
COPD	0.92	0.52, 1.53	0.754	0.90	0.51, 1.51	0.702	0.91	0.52, 1.53	0.737
BMI (kg/m^2)	0.98	0.95, 1.01	0.302	0.98	0.95, 1.01	0.209	0.98	0.95, 1.01	0.128
Hypertension	0.96	0.67, 1.38	0.819	0.96	0.67, 1.38	0.827	0.93	0.64, 1.34	0.683
Coronary artery disease	0.77	0.42, 1.35	0.383	0.77	0.42, 1.34	0.375	0.66	0.36, 1.17	0.170
Cerebrovascular disease	1.47	0.63, 3.12	0.340	1.46	0.62, 3.09	0.352	1.51	0.64, 3.23	0.312
Heart failure	0.37	0.02, 2.02	0.352	0.36	0.02, 1.96	0.336	0.41	0.02, 2.27	0.409
High risk surgery	3.42	2.37, 5.02	<0.001	3.41	2.36, 5.00	<0.001	3.31	2.28, 4.88	<0.001
Orthopedic surgery	0.39	0.18, 0.76	0.009	0.39	0.18, 0.76	0.009	0.42	0.19, 0.81	0.016
Low non-sex-specific peak VO_2[a]				1.41	0.89, 2.20	0.136			
Low sex-specific peak VO_2[b]							2.29	1.60, 3.30	<0.001

95% confidence intervals and p-values of predictor variables were obtained by Wald's test. [a]14 mL kg^{-1} min^{-1}. [b]Males: 22.3 mL kg^{-1} min^{-1}; Females: 12.4 mL kg^{-1} min^{-1}; OR: Odds Ratio; CI: Confidence interval.

Table 4: **Multivariable logistic regression models for the association of low peak VO_2 with moderate-to-severe postoperative complications.**

However, the net reclassification improvement by the sex-specific model was substantially higher, improving classification for 39% of patients (event NRI = 0.29, non-event NRI = 0.1, overall NRI = 0.39, 95% CI for overall NRI 0.24, 0.55). The Hosmer–Lemeshow test showed a good fit for both the non-sex-specific model (p = 0.9996) and the sex-specific model (p = 0.9114). Additionally, calibration plots revealed a good calibration as shown by a bias-corrected calibration curve (Figure A-1, Appendix).

Sex-specific association of AT and peak VO_2 with DASI

The mean (SD) DASI score was lower in females than males (38.4 (14.4) versus 45 (13.7); p < 0.001), respectively. A linear regression analysis was performed with either peak VO_2 or AT as the dependent variable and with an interaction term between DASI score and sex as independent variables (Table A-1). Peak VO_2 positively correlated with DASI in both males and females (Fig. 1A). A significant interaction with a positive coefficient of the interaction term (coefficient for male sex 0.06, p = 0.019) was observed, with a significantly steeper positive correlation between DASI score and peak VO_2 in males than females. Furthermore, peak VO_2 estimated through DASI was higher than the actual CPET-derived peak VO_2 in both sexes, with females being overestimated both to a greater extent (9.3 versus 7.7 mL kg^{-1} min^{-1}; p < 0.001) and more frequent (86.5 versus 83.5%; p = 0.0021). Furthermore, the correlation between DASI and AT was investigated. No significant interaction was observed between sex and DASI score in the linear regression analysis (Fig. 1A, Table A-1).

Discussion

Cardiopulmonary exercise testing (CPET) has an established track record of objectively assessing the functional capacity of patients, and increasingly so for surgical risk assessment and to guide exercise-based prehabilitation strategies. However, sex-specific aspects remain overlooked in interpreting CPET-derived data to date. In the original METS study, among all CPET parameters, peak VO_2 was most strongly associated with moderate-to-severe postoperative complications, and hence we investigated this association separately for both biological sexes. Within the study population, more females than males were below the established non-sex-specific peak VO_2 threshold of 14 mL kg^{-1} min^{-1} (4 METS), a threshold that is widely accepted for preoperative risk assessment.[23] Males exhibited a higher rate of postoperative complications than females with similar risk profiles despite higher preoperative peak VO_2 values, suggesting a higher functional capacity. This highlights the need for a sex-specific approach to

than females. B: Plots of probability density function (left panel) and cumulative distribution function (right panel) stratified by postoperative complications (POC) and biological sex. The black dashed line represents the non-sex-specific peak VO_2 threshold of 14 mL kg^{-1} min^{-1}, while the blue and red dashed line represent the optimal peak VO_2 threshold for males and females, respectively. C: The optimal threshold for surgical risk prediction was determined via maximization of the Youden index. For males, a relatively broad interval yields approximately equal Youden indices. At 22.3 mL kg^{-1} min^{-1}, the Youden index is maximal for males. A correspondingly lower threshold of 12.4 mL kg^{-1} min^{-1} was obtained for females. AT, Anaerobic threshold; DASI, Duke Activity Status Index; VO_2, oxygen consumption.

functional capacity parameters as reliable predictors of postoperative outcomes.

We could identify sex-specific peak VO_2 thresholds and apply them in a classification model for moderate-to-severe complications. More males than females were ranked below their corresponding sex-specific threshold, translating into poorer functional capacity and thus potentially higher surgical risk. This poses a possible explanation for the higher complication rate in males, as their functional capacity relative to surgical risk may subsequently be overestimated by the sex-unspecific peak VO_2 threshold, whereas females appear to be underestimated. Indeed, sex-specific threshold values were more accurate in their association with moderate-to-severe complications. We demonstrated this by comparing logistic regression models with low peak VO_2 as a feature variable either defined by the sex-unspecific threshold or the computed sex-specific threshold. Measures of discriminatory ability such as AUC-ROC and NRI were higher in the sex-specific model. Applying the sex-specific peak VO_2 led to a higher proportion (39% of the study cohort) of study patients being more correctly classified compared to the baseline model. In contrast, the sex-unspecific model only led to an improvement in risk classification of 9% of patients.

Besides objective assessment of functional capacity we also investigated sex-related differences of the DASI questionnaire, which is widely used in the Anglo-American region. DASI allows an approximation of peak VO_2 based on subjective assessment of functional capacity by a 12-item questionnaire about daily activities.[28] We compared the DASI-derived approximated peak VO_2 with the objectively identified values obtained by CPET and could demonstrate that DASI overestimated peak VO_2 in females more often and to a greater extent than it did in males. The linear relationship of DASI score and true peak VO_2 is modulated by patient's sex as revealed by a statistical interaction within linear regression analysis. In males, there is a steeper correlation between DASI and peak VO_2 than in females. Thus, in males, actual peak VO_2 increases more strongly with DASI, suggesting a more substantial impact of subjective functional capacity on its objective counterpart. However, this could be potentially rooted in musculoskeletal differences that generate performance differences in the DASI questionnaire regardless of cardiopulmonary fitness.

The appropriate consideration of sex in the medical field is increasingly scrutinized. A review by Mauvais-Jarvis et al.[37] highlights the kaleidoscope of factors influencing health, disease, and medicine. Sex is a genetic modifier of biology and disease and carries significant implications for sex-specific precision medicine, necessitating sex-specific medical practice recommendations. Similarly, diagnostic tools such as CPET should be tailored to the patient's sex, and CPET parameters should be adjusted as a basis for perioperative risk prediction and clinical decision-making. CPET-derived values are embedded into clinical practice and define risk for surgery, used to defer surgery for prehabilitation purposes or even recommend against surgery in deconditioned patients. In a comprehensive data analysis of the FRIEND registry, Kaminsky et al.[38] found significant sex-related differences in peak VO_2, which aligns with our findings. Indeed, sex-specific CPET differences are found in athletes,[39,40] cancer patients,[41,42] and other patient groups.[43,44] However, these differences have not yet been accounted for when using CPET for surgical risk prediction. Thomas et al.[20] similarly reported sex-related differences in a *post hoc* analysis of 703 patients in the perioperative setting with AT and peak VO_2 values being significantly higher in males than females before and after correction for body weight. However, it is important to note that biological sex can influence multiple aspects of health, but the existence of gender-related aspects must not be disregarded. Gender-related lifestyle and behavioral diversity may also have influenced the observed differences in biological sex.

Despite our consistent findings, there are several limitations of the study that must be considered. Since this work is a *post hoc* analysis of a prospective study, the primary endpoint in this work differs from that the original METS study, where postoperative complications were considered a secondary endpoint. This *post hoc* choice of endpoint must be considered a methodological limitation. Furthermore, the consent rate of eligible patients of the original METS study was only 20%, with potential of a selection-bias. Yet, it is known from previous CPET trials that a dropout rate of 10–20% is not unusual and oftentimes related to the busy clinical schedule that patients have while being prepared for surgery and while in the hospital.[45,46] Of those patients that consented to the METS study, CPET was not performed or had incomplete data in 145 patients, with the potential that patients with higher risk were excluded from our *post hoc* analysis. In consideration of this, we evaluated whether these patients who underwent surgery without preoperative CPET had a higher complication rate than the cohort from which we performed our *post hoc* analysis. Reassuringly, in the cohort of patients that were excluded due to absence of CPET data, 10.3% had moderate-to-severe complications. The complication rate in the analyzed cohort was even higher (13.9%), although the difference was not statistically significant via the chi-squared test ($p = 0.29$). Consequently, it is unlikely that the exclusion of a small cohort of patients biased the study results. On the other hand, we found indeed some differences regarding comorbidities. The excluded cohort ($n = 145$) showed a higher rate of diabetes mellitus (25% versus 18%, $p = 0.04$), and higher rate of heart failure (4.1% versus 1.4%, $p = 0.03$). Moreover, DASI was significantly lower in the excluded individuals (32.6 (16.1) versus 42.5 (14.3), $p < 0.001$), suggesting that in certain aspects sicker patients did not undergo CPET.

In the original METS study, where myocardial infarction was a primary endpoint, its incidence was lower than expected. Indeed, in our study cohort the incidence of perioperative myocardial infarction and myocardial injury was low with an incidence of 1.4% and 11.7%, respectively. In comparison, Puelacher et al.[47] reported 4.6% and 16% incidence of perioperative myocardial infarction and myocardial injury, respectively.

Other limitations include the fact that the METS study had an imbalance of enrollment of males and females, with twice as many males included in the study compared to females. As such, our study cohort is not fully representative for the noncardiac surgical population with regards to sex distribution. However, we know from previous exercise studies that female patients are willing to exercise preoperatively in a similar manner to male patients.[20,48] Additionally, our analysis does not consider the fact that body composition differs between males and females, and the fact that differences in lean body mass may have influenced our results. However, data on this is unfortunately not available. Similarly, differences between males and females may not only be related to differences in functional capacity related to biological sex. It is important to note that biological sex can also influence multiple aspects of health, including immune responses, hormonal profiles and certainly gender-related aspects. Determinants related to gender, including healthcare-seeking behaviors, and adherence to care/research could also have played a role in patient enrollment. This may align with our findings for the relationship between DASI and peak VO_2, where self-declared subjective assessment of functional capacity might be affected by sex- and gender-related aspects, which might cause overestimation of peak VO_2 by females compared to males. Given that females received orthopedic surgery significantly more frequent in this cohort, it is possible that limitations in their capacity to perform CPET due to joint pain and mobility issues accounted for lower average peak VO_2, rather than solely reflecting differences in cardiovascular functional capacity. Thus, we adjusted the classification model for orthopedic procedures besides to high-risk surgery and found that orthopedic procedures were associated with a lower risk of moderate-to-severe complications.

Regarding the potential biological pathway between functional capacity and complications of infectious or respiratory origin, we acknowledge that further investigation is required to elucidate the precise mechanisms involved. Individuals with higher functional capacity often exhibit better overall physiological reserve, including immune function, which could potentially contribute to a reduced risk of postoperative infections. Additionally, higher functional capacity is associated with better oxygenation and ventilation during physical exertion, potentially minimizing the risk of postoperative respiratory, perfusion, and wound complications. We acknowledge the need for additional research to better understand the specific mechanisms involved, as well as the potential reasons for the observed differences between males and females.

Evaluation of model performance was assessed by quantifying AUC-ROC. In our case, the sex-specific model had a statistically significant higher AUC-ROC, as determined per DeLong's test, and was 0.73 compared to 0.7. While this may not appear a numerically large increase it is important to remember that the increase in AUC often has a small magnitude in nested logistic regression models. This is because the AUC calculates the average sensitivity over all calculated values of specificity against all theoretical thresholds of the test, thereby also considering clinically irrelevant thresholds. Thus, changes in clinically relevant threshold ranges may be faded by inverse or absent effects in clinically irrelevant threshold ranges. It is known that only very strong associations of a new predictor will noticeably increase the AUC-ROC, to the point that the new predictor is associated with an exceptionally high odds ratio.[49–51] For this reason, we decided to use the NRI alongside the AUC as an additional estimate for discriminatory power, which showed a meaningful improvement in the discrimination of the sex-specific model. Importantly, while the establishment of threshold values for continuous variables is preferable for clinical routine to allow pragmatic classification into, for example, low risk versus high risk groups, it should be remembered that dichotomization of a continuous variable is accompanied by relevant information loss. Therefore, algorithms that implement continuous peak VO_2 as a sex-specific feature and estimate perioperative risk suggest a potential future approach.

Our *post hoc* analysis of study data from 1266 patients who underwent CPET and a DASI survey before surgery showed that despite similar risk profiles and comorbidities, CPET parameters differ according to sex. It is reasonable to suppose different reference values for males and females with respect to the distributions of peak VO_2 among sex. Consequently, there are implications for the risk stratification of postoperative complications based on peak VO_2. We were able to identify sex-specific threshold values and compared these with a widely accepted non-sex-specific threshold value within classification models. Here, the sex-specific model was able to dominate in terms of its discriminatory ability. Since biological sex and perhaps also gender-associated factors potentially interfere with the relationship between peak VO_2 and postoperative complications, it is suggested to consider applying sex-specific threshold values when evaluating CPET examinations in the context of preoperative exercise testing for clinical decision-making, e.g., referral for prehabilitation or deferment of surgery, lest we perpetuate disparity in healthcare.

Contributors
JA, BR, HI, DNW, BHC and RS contributed to the conception and design of the study. JA, BR, HI, KH, SX, DNW, BHC and RS contributed to the acquisition, analysis, and interpretation of the data. JA and RS wrote the first draft of the manuscript. JA, BR, HI, KH, PZ, TK, DNW and BHC revised the manuscript critically. All authors read and approved the final version of the manuscript. JA, BR and RS were responsible for the decision to submit the manuscript.

Data sharing statement
Patient data originates from the METS study and is not publicly shared outside the METS study group. The statistical analysis plan of the present study will be made available upon reasonable request from the corresponding author after a signed data access agreement.

Declaration of interests
All authors declare no competing interests.

Acknowledgements
This study was conducted on behalf of the METS study group. The original METS study received funding for St. Michael's Hospital (Toronto, Ontario, Canada) from Canadian Institutes of Health Research, from Heart and Stroke Foundation of Canada, and from Ontario Ministry of Research, Innovation and Science. University Health Network (Toronto, Ontario, Canada) received funding from the Ontario Ministry of Health and Long-Term Care. Queen Mary University of London, UK received funding from UK National Institute of Academic Anaesthesia and from UK Clinical Research Collaboration. Monash University (Melbourne, Victoria, Australia) received funding from Australian and New Zealand College of Anaesthetists, and Monash University.

Appendix A. Supplementary data
Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2023.102223.

References

1 Nepogodiev D, Martin J, Biccard B, Makupe A, Bhangu A, National Institute for Health Research Global Health Research Unit on Global Surgery. Global burden of postoperative death. *Lancet*. 2019;393:401.

2 Wijeysundera DN, Pearse RM, Shulman MA, et al. Assessment of functional capacity before major non-cardiac surgery: an international, prospective cohort study. *Lancet*. 2018;391:2631–2640.

3 Levett DZH, Jack S, Swart M, et al. Perioperative cardiopulmonary exercise testing (CPET): consensus clinical guidelines on indications, organization, conduct, and physiological interpretation. *Br J Anaesth*. 2018;120:484–500.

4 Older P, Smith R, Courtney P, Hone R. Preoperative evaluation of cardiac failure and ischemia in elderly patients by cardiopulmonary exercise testing. *Chest*. 1993;104:701–704.

5 Older P, Hall A, Hader R. Cardiopulmonary exercise testing as a screening test for perioperative management of major surgery in the elderly. *Chest*. 1999;116:355–362.

6 Snowden CP, Prentis JM, Anderson HL, et al. Submaximal cardiopulmonary exercise testing predicts complications and hospital length of stay in patients undergoing major elective surgery. *Ann Surg*. 2010;251:535–541.

7 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;112:491–497.

8 Colson M, Baglin J, Bolsin S, Grocott MPW. Cardiopulmonary exercise testing predicts 5 yr survival after major surgery. *Br J Anaesth*. 2012;109:735–741.

9 Hightower CE, Riedel BJ, Feig BW, et al. A pilot study evaluating predictors of postoperative outcomes after major abdominal surgery: physiological capacity compared with the ASA physical status classification system. *Br J Anaesth*. 2010;104:465–471.

10 Wilson RJT, Davies S, Yates D, Redman J, Stone M. Impaired functional capacity is associated with all-cause mortality after major elective intra-abdominal surgery. *Br J Anaesth*. 2010;105:297–303.

11 West MA, Lythgoe D, Barben CP, et al. Cardiopulmonary exercise variables are associated with postoperative morbidity after major colonic surgery: a prospective blinded observational study. *Br J Anaesth*. 2014;112:665–671.

12 Junejo MA, Mason JM, Sheen AJ, et al. Cardiopulmonary exercise testing for preoperative risk assessment before hepatic resection. *Br J Surg*. 2012;99:1097–1104.

13 Neviere R, Edme JL, Montaigne D, Boleslawski E, Pruvot FR, Dharancy S. Prognostic implications of preoperative aerobic capacity and exercise oscillatory ventilation after liver transplantation. *Am J Transplant*. 2014;14:88–95.

14 Tolchard S, Angell J, Pyke M, et al. Cardiopulmonary reserve as determined by cardiopulmonary exercise testing correlates with length of stay and predicts complications after radical cystectomy. *BJU Int*. 2015;115:554–561.

15 Brunelli A, Belardinelli R, Pompili C, et al. Minute ventilation-to-carbon dioxide output (VE/VCO2) slope is the strongest predictor of respiratory complications and death after pulmonary resection. *Ann Thorac Surg*. 2012;93:1802–1806.

16 Carlisle JB, Danjoux G, Kerr K, Snowden C, Swart M. Validation of long-term survival prediction for scheduled abdominal aortic aneurysm repair with an independent calculator using only pre-operative variables. *Anaesthesia*. 2015;70:654–665.

17 Grant SW, Hickey GL, Wisely NA, et al. Cardiopulmonary exercise testing and survival after elective abdominal aortic aneurysm repair†. *Br J Anaesth*. 2015;114:430–436.

18 Snowden CP, Prentis J, Jacques B, et al. Cardiorespiratory fitness predicts mortality and hospital length of stay after major elective surgery in older people. *Ann Surg*. 2013;257:999–1004.

19 Daugherty SL, Magid DJ, Kikla JR, et al. Gender differences in the prognostic value of exercise treadmill test characteristics. *Am Heart J*. 2011;161:908–914.

20 Thomas G, West MA, Browning M, et al. Why women are not small men: sex-related differences in perioperative cardiopulmonary exercise testing. *Perioper Med (Lond)*. 2020;9:18.

21 Molenaar CJL, Minnella EM, Coca-Martinez M, et al. Effect of multimodal prehabilitation on reducing postoperative complications and enhancing functional capacity following colorectal cancer surgery: the PREHAB randomized clinical trial. *JAMA Surg*. 2023;158:572–581.

22 Waterland JL, McCourt O, Edbrooke L, et al. Efficacy of prehabilitation including exercise on postoperative outcomes following abdominal cancer surgery: a systematic review and meta-analysis. *Front Surg*. 2021;8:628848.

23 Fleisher LA, Fleischmann KE, Auerbach AD, et al. 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery. *Circulation*. 2014;130:2215–2245.

24 Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J*. 2016;37:2129–2200.

25 Yancy CW, Jessup M, Bozkurt B, et al. 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the american college of cardiology/american heart association task force on clinical practice guidelines and the heart failure society of America. *Circulation*. 2017;136:e137–e161.

26 American Thoracic Society and American College of Chest Physicians. ATS/ACCP statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med*. 2003;167:211–277.

27 Gaskill SE, Ruby BC, Walker AJ, Sanchez OA, Serfass RC, Leon AS. Validity and reliability of combining three methods to determine ventilatory threshold. *Med Sci Sports Exerc*. 2001;33:1841–1848.

28 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;64:651–654.

29 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;116:177–191.

30 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;103:744–752.
31 Jammer I, Wickboldt N, Sander M, et al. Standards for definitions and use of outcome measures for clinical effectiveness research in perioperative medicine: European perioperative clinical outcome (EPCO) definitions: a statement from the ESA-ESICM joint task-force on perioperative outcome measures. *Eur J Anaesthesiol*. 2015;32:88–105.
32 International Surgical Outcomes Study (ISOS) group. Prospective observational cohort study on grading the severity of postoperative complications in global surgery research. *Br J Surg*. 2019;106:e73–e80.
33 Thygesen K, Alpert JS, Jaffe AS, et al. Third universal definition of myocardial infarction. *Circulation*. 2012;126:2020–2035.
34 Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32–35.
35 R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2022. https://www.R-project.org/.
36 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.
37 Mauvais-Jarvis F, Bairey Merz N, Barnes PJ, et al. Sex and gender: modifiers of health, disease, and medicine. *Lancet*. 2020;396:565–582.
38 Kaminsky LA, Imboden MT, Arena R, Myers J. Reference standards for cardiorespiratory fitness measured with cardiopulmonary exercise testing using cycle ergometry: data from the fitness registry and the importance of exercise national database (FRIEND) registry. *Mayo Clin Proc*. 2017;92:228–233.
39 Parsonage JR, Secomb JL, Tran TT, et al. Gender differences in physical performance characteristics of elite surfers. *J Strength Cond Res*. 2017;31:2417–2422.
40 Henriques-Neto D, Magalhães JP, Hetherington-Rauth M, Santos DA, Baptista F, Sardinha LB. Physical fitness and bone health in young athletes and nonathletes. *Sports Health*. 2020;12:441–448.
41 Jee H, Kim J-H. Gender difference in colorectal cancer indicators for exercise interventions: the national health insurance sharing service-derived big data analysis. *J Exerc Rehabil*. 2019;15:811–818.
42 Titz C, Hummler S, Schmidt ME, Thomas M, Steins M, Wiskemann J. Exercise behavior and physical fitness in patients with advanced lung cancer. *Support Care Cancer*. 2018;26:2725–2736.
43 Ferns SJ, Wehrmacher WH, Serratto M. Effects of obesity and gender on exercise capacity in urban children. *Gend Med*. 2011;8:224–230.
44 Salvioni E, Corrà U, Piepoli M, et al. Gender and age normalization and ventilation efficiency during exercise in heart failure with reduced ejection fraction. *ESC Heart Fail*. 2020;7:371–380.
45 Schenk A, Pulverer W, Koliamitra C, et al. Acute exercise increases the expression of KIR2DS4 by promoter demethylation in NK cells. *Int J Sports Med*. 2019;40:62–70.
46 Schenk A, Koliamitra C, Bauer CJ, et al. Impact of acute aerobic exercise on genome-wide DNA-methylation in natural killer cells-a pilot study. *Genes*. 2019;10:380.
47 Puelacher C, Lurati Buse G, Seeberger D, et al. Perioperative myocardial injury after noncardiac surgery: incidence, mortality, and characterization. *Circulation*. 2018;137:1221–1232.
48 Bauer CJ, Findlay M, Koliamitra C, et al. Preoperative exercise induces endothelial progenitor cell mobilisation in patients undergoing major surgery - a prospective randomised controlled clinical proof-of-concept trial. *Heliyon*. 2022;8:e10705.
49 Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol*. 2004;159:882–890.
50 Ware JH. The limitations of risk factors as prognostic tools. *N Engl J Med*. 2006;355:2615–2617.
51 Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27:157–172. discussion 207-12.