Langenbeck's Archives of Surgery (2024) 409:372
https://doi.org/10.1007/s00423-024-03560-0

RESEARCH

Patient-reported questionnaires to preoperatively identify high-risk surgical patients

Renske Meijer[1,2] · David W. G. ten Cate[3,4] · Bart C. Bongers[2,4] · Marta Regis[5] · Hans H. C. M. Savelberg[2] · Gerrit D. Slooter[3] · Stef Janssen[6] · Martijn van Hooff[1,2] · Goof Schep[1]

Received: 8 August 2024 / Accepted: 22 November 2024

Abstract

Purpose Low cardiorespiratory fitness (CRF) increases the risk of postoperative morbidity and mortality following major surgery. Assessing CRF preoperatively, by measuring peak oxygen uptake (VO_{2peak}) during cardiopulmonary exercise testing (CPET), is valuable yet not widely available. This study aimed to assess whether questionnaires could be used preoperatively to identify high-risk surgical patients.

Methods Healthy participants and patients who underwent CPET completed the FitMáx, Duke Activity Status Index (DASI), the modified 4-questions DASI (M-DASI-4Q), Veterans-Specific Activity Questionnaire (VSAQ), and Metabolic Equivalents of Task (MET) questionnaire. Questionnaire-VO_{2peak} was compared with CPET-VO_{2peak}. Overall performance of the questionnaires was assessed by the area under the curve (AUC) of receiver operating characteristic (ROC) curves. Furthermore, corresponding to the Youden index or pre-specified levels, sensitivity, specificity, and predictive values were determined.

Results In total, 361 participants were included. All questionnaires showed high AUC values to identify high-risk patients, defined on the basis of CPET-VO_{2peak} thresholds. FitMáx and VSAQ demonstrated superior results compared to the other questionnaires. Based on the Youden index, the optimal questionnaire-VO_{2peak} cut-off values were 20.6, 21.3, and 26.1 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$ for the FitMáx and 16.3, 18.2, and 20.4 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$ for the VSAQ corresponding to the VO_{2peak} thresholds 16.0, 18.2 and 24.5 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$ respectively.

Conclusion The ability to identify high-risk surgical patients preoperatively (defined by the CPET-VO_{2peak} thresholds) by the FitMáx and the VSAQ indicates that they could be used to identify high-risk surgical patients. Patients with a poor predicted $VO_{2peak} \leq 21.3$ and ≤ 18.2 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$, respectively for FitMáx and VSAQ, should be referred to formal preoperative (cardiopulmonary) exercise testing.

Trial registration The study was registered as NL-OMON23304 in the Overview of Medical Research in the Netherlands, retrospectively at 28–04-2020.

Keywords Aerobic fitness · Cardiopulmonary exercise testing · Preoperative risk assessment · Preoperative care · Patient-reported outcome measure

✉ Renske Meijer
renske.meijer@mmc.nl

1 Department of Sports Medicine, Máxima Medical Centre, De Run 4600, 5504 DB Veldhoven, The Netherlands

2 Department of Nutrition and Movement Sciences, NUTRIM, Institute of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, The Netherlands

3 Department of Surgery, Máxima Medical Centre, Veldhoven, The Netherlands

4 Department of Surgery, NUTRIM, Institute of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, The Netherlands

5 Department of Mathematics and Computer Science, University of Technology, Eindhoven, The Netherlands

6 Department of Anaesthesiology, Máxima Medical Centre, Veldhoven, The Netherlands

Published online: 05 December 2024

Introduction

Cardiorespiratory fitness (CRF) is defined as the integrated capacity to take in, transport, and utilise oxygen; thus it provides a reflection of an individual's functional capacity [1]. Surgery imposes increased metabolic demands on the body's cardiorespiratory system. Patients with low CRF are more prone to worse surgical outcomes, such as complications, longer length of hospital stay, and postoperative mortality [2–4]. Especially in high-risk patients, promising effects of (p)rehabilitation programmes are observed, emphasising the importance to timely identify and counsel patients with low CRF to consider preventive preoperative interventions [5–9].

In current clinical preoperative practice, functional capacity is often assessed subjectively by anaesthesiologists, using a scoring system such as the American Society of Anaesthesiologists (ASA) physical status classification [10, 11]. In contrast, CRF assessment by objectively measuring oxygen uptake at peak exercise (VO_{2peak}) during maximal cardiopulmonary exercise testing (CPET) is considered the gold standard and leads to a more accurate risk assessment [11, 12]. CPET provides useful information for perioperative risks and detects the cause of exercise limitation [13, 14]. Several risk thresholds have been described for objectively measured preoperative VO_{2peak} to predict postoperative complications [15–17]. However, performing CPET is costly, labour-intensive, and interpretation is difficult, especially in patients with (multiple) comorbidities; therefore, it is not widely applicable for preoperative risk assessment [18, 19]. Moreover, CPET outcome measures showed a consistent variability in repeated testing in unfit patients, leading to the requirement for more reliable ways to determine VO_{2peak} [20].

Patient-reported questionnaires are frequently used as an alternative to estimate VO_{2peak} [21]. Examples of such questionnaires are the FitMáx© questionnaire (hereafter FitMáx) [22], the Duke Activity Status Index (DASI) [23], the modified version of the DASI with only four questions (M-DASI-4Q) [24], the Veterans-Specific Activity Questionnaire (VSAQ) [25], and the Metabolic Equivalents of Task (MET) questionnaire (commonly used in The Netherlands) based on the international compendium for MET [26]. A limitation of these questionnaires is the use of activities that are not performed worldwide (e.g., basketball, cross-country skiing, cycling) [22, 27]. Correlations of these self-reported questionnaires with objectively assessed CPET-VO_{2peak} vary among different study populations, and particularly detecting patients with a low CRF remains difficult [12, 23, 25, 28–31]. The accuracy of these questionnaires to detect predefined clinically relevant VO_{2peak} thresholds is unknown. Therefore, the main aim of this study was to evaluate whether self-reported questionnaires could be used in the preoperative evaluation to identify high-risk surgical patients, using predefined clinically relevant VO_{2peak} thresholds corresponding to a higher risk for complications.

Material and methods

Setting and population

Patients and healthy participants aged ≥ 18 years who were scheduled for CPET, either for medical purposes or as part of a general health check, were approached for participation in the current study. Data was collected at the Máxima Medical Centre (MMC), a Dutch non-academic teaching hospital, and Ancora Health, a digital start-up in preventive healthcare in the Netherlands, from 2018–2024. Ethical approval for this study (reference number N18.051) was provided by the authorised Medical Research Ethics Committee of the MMC, Veldhoven, on 30 April 2018 (reference number N18.051). Data from participants in previous (external validation group of the FitMáx prediction model) and ongoing studies were included retrospectively in the current study [22, 30, 32]. Subjects were eligible for inclusion if they were proficient in the Dutch language and completed the questionnaires within 42 days from the day of CPET. The ASA classification, of the included general population, was determined following the guidelines of the American society of Anaesthesiologists to assess the participants' physical status [10, 11]. All participants provided written informed consent for the use of their data.

Cardiopulmonary exercise testing

CPET was conducted in accordance with international recommendations [33]. Pre-test instructions included intake of regular medication, as well as no heavy exercise and no large meal two hours prior to the exercise test. The CPET protocol is extensively described in our FitMáx validation study [22]. All tests aborted for medical reasons (substantial ECG changes, cardiac symptoms, severe orthopaedic problems) were excluded. VO_{2peak} was calculated using the 30-s average of VO_2 values prior to peak exercise and was subsequently normalised for body mass.

Self-reported questionnaires

FitMáx consists of three single-answer questions to estimate a subject's maximum capacity of walking, cycling, and stair climbing on a 14-point, 12-point, and 11-point scale, respectively. FitMáx scores for each item were incorporated with age, sex, and body mass index into a prediction algorithm to estimate VO_{2peak} [22]. The DASI includes 12 daily activities, of which a weighted sum score leads to an estimated VO_{2peak} value. Scores range from 0 to 58.2, resulting in VO_{2peak} values from 9.6 to 34.6 ml·kg^{-1}·min^{-1} [23]. More recently, a study on the M-DASI-4Q was performed, in which only four questions of the original DASI were selected to predict the preoperative risk of complications [24]. For the M-DASI-4Q, no algorithm to estimate VO_{2peak} was available. As such, only the score of the M-DASI-4Q on a scale from 1–4 was included. The VSAQ focuses on daily activities ordered with increasing difficulty on a 13-point scale, of which the first limiting activity corresponds to an equivalent number of maximal METs. This MET score is combined with age to estimate VO_{2peak} [25]. The results of the FitMáx, DASI, and VSAQ were converted into VO_{2peak} in ml·kg^{-1}·min^{-1}, following the guidelines of each questionnaire [22, 23, 25]. The MET questionnaire is applied preoperatively by anaesthesiologists in (Dutch) hospitals, to assess CRF. It consists of 9 items with increasing MET values based on the globally used MET compendium of physical activities, in which the subject is asked to rate their maximum capacity [26]. VO_{2peak} values were calculated by multiplying the reported MET scores by 3.5 ml·kg^{-1}·min^{-1}.

Statistical analysis

Normality of the data was visually evaluated by histograms and naïve assessment of skewness and kurtosis [34]. Continuous variables are reported as mean ± standard deviation in case of normal distribution, and median and interquartile range (IQR) otherwise. Categorical variables are presented as count *n* with percentage (%).

The American Heart Association stated that for patients with a MET value ≥ 7 (VO_{2peak} ≥ 24.5 ml·kg^{-1}·min^{-1}), preoperative CPET had no additional value, since the risk of perioperative cardiovascular events was low [17]. Furthermore, literature suggests that a VO_{2peak} of ≤ 16.0 or ≤ 18.2 ml·kg^{-1}·min^{-1} is significantly associated with an increased risk of postoperative complications [15, 16]. These VO_{2peak} thresholds were selected for the current study as an indicator of a higher preoperative risk for complications. Receiver operating characteristic (ROC) curves were used as a graphical approach to compare the performance of the questionnaires, separately at the considered VO_{2peak} thresholds (≤ 16.0, ≤ 18.2, and ≤ 24.5 ml·kg^{-1}·min^{-1}). The area under the curve (AUC) was determined, with corresponding 95% confidence intervals (CI) [35]. Also, the optimal cut-off value in the prediction from the questionnaires was determined as the one maximizing the Youden index, the maximum difference between sensitivity (the probability of correctly classifying subjects below the predefined VO_{2peak} thresholds), and the complementary of specificity (the probability of incorrectly classifying subjects above the predefined VO_{2peak} thresholds) [36]. In clinical practice, a high sensitivity may be more important than specificity to prevent false negative outcomes [37]. Since the Youden index does not distinguish between the importance of these two measures, the optimal FitMáx cut-off value for predetermined sensitivity values from 0 to 1 in steps of 0.05 were also presented [38]. Sensitivity, specificity and predictive values were calculated for the questionnaire cut-off values [37, 39]. We have also quantitatively compared the considered metrics (AUC [35], Youden index [36, 40], and predetermined sensitivity [38]), pairwise between the FitMáx and the other questionnaires, and applied the Benjamini–Hochberg correction for multiple testing [41]. This was done in the pROC package in R, using the stratified bootstrap option (10000 bootstrap replicates; retaining similar case-controls observations) [35, 42, 43]. To assess the performance of the questionnaires in a subpopulation that is at higher preoperative risk for complications (ASA > II) [11], the same analyses were repeated on two subgroups based on the ASA physical status classification (I-II and III-IV). Statistical analysis was performed using R (R-version 4.3.1) [44]. All p-values < 0.05 were considered statistically significant.

Results

In total, 361 participants (277 men and 84 women) were eligible for analyses (e.g., complete cases, within 42 days of CPET). Participants' age ranged from 19−90 years with a VO_{2peak} ranging from 7.8−71.4 ml•kg^{-1}•min^{-1}. Cardiac evaluation (58%) was the most common reason for CPET in men and pulmonary evaluation (55%) was the most common reason in women, in the current study population. The most frequent physical status classification was ASA III (32%) (Table 1). Median (IQR) CPET-VO_{2peak} was 27.6 (17.0–40.0) ml·kg^{-1}·min^{-1} for men, and 19.5 (16.0–26.6) ml·kg^{-1}·min^{-1} for women, reflecting on average 91% and 100% of the reference value for healthy Dutch males and females, respectively [45].

The AUC of the FitMáx was significantly higher compared to the DASI, MET-questionnaire, and the M-DASI-4Q for the included thresholds, in the total study population (Fig. 1a-c). An AUC of 0.93 (0.90–0.95), 0.93 (0.90–0.95), and 0.97 (0.95–0.99) was found to detect patients at risk based on the VO_{2peak} ≤ 16.0, ≤ 18.2, and ≤ 24.5 ml·kg^{-1}·min^{-1}, respectively. The AUC of the FitMáx was not significantly different compared to the VSAQ, except for the predefined VO_{2peak} threshold of ≤ 24.5 ml·kg^{-1}·min^{-1}, in which a statistically significant difference was found.

The FitMáx cut-off values corresponding to the maximal value of the Youden index, obtained high sensitivity

Table 1 Participant baseline characteristics

Variable	Total (100%, $n=361$)	Male (77%, $n=277$)	Female (23%, $n=84$)
n			
Anthropometrical data			
Age (years) *	61 [51–70]	61 [52–70]	58 [46–68]
Body height (cm) *	176 [169–183]	179 [174–184]	165 [161–171]
Body mass (kg) *	81 [72–92]	83 [76–93]	71 [61–83]
BMI (kg•m^{-2})	25.9 [23.6–29.0]	25.8 [23.9–28.9]	26.0 [22.1–29.1]
FEV_1 (L) *	3.24 [2.36–3.95]	3.49 [2.70–4.11]	2.25 [1.71–2.97]
FVC (L) *	4.10 [3.18–4.96]	4.44 [3.63–5.10]	2.91 [2.40–3.66]
COPD, GOLD classification			
None	325 (90%)	250 (90%)	75 (89%)
GOLD I	5 (1%)	3 (1%)	2 (2%)
GOLD II	17 (5%)	12 (4%)	5 (6%)
GOLD III	11 (3%)	10 (4%)	1 (1%)
GOLD IV	3 (1%)	2 (1%)	1 (1%)
Use of β-blocker			
Yes *	115 (32%)	97 (35%)	18 (21%)
No	246 (68%)	180 (65%)	66 (79%)
ASA-Classification *			
Missing	54 (15%)	48 (17%)	6 (7%)
I	66 (18%)	45 (16%)	21 (25%)
II	85 (24%)	59 (21%)	26 (31%)
III	117 (32%)	89 (32%)	28 (33%)
IV	39 (11%)	36 (13%)	3 (4%)
Reason CPET *			
Health check-up	59 (16%)	50 (18%)	9 (11%)
Cardiac evaluation *	183 (51%)	160 (58%)	23 (27%)
Pulmonary evaluation *	103 (29%)	57 (21%)	46 (55%)
Oncologic rehabilitation	16 (4%)	10 (4%)	6 (7%)
CPET data			
VO_{2peak} (ml•kg^{-1}•min^{-1}) *	25.3 [17.2–38.4]	27.6 [17.0–40.0]	19.5 [16.0–26.6]
WR_{peak} (W) *	176 [102–285]	208 [113–315]	108 [80–160]
HR_{peak} (beats•min^{-1})	155 [131–171]	157 [131–171]	153 [131–169]
RER_{peak}	1.14 [1.08–1.20]	1.15 [1.09–1.21]	1.12 [1.07–1.18]
Exercise time (min)	9.3 [8.2–10.4]	9.4 [8.4–10.4]	8.9 [7.6–10.4]
VO_{2peak} reference [a] (ml•kg^{-1}•min^{-1}) *	29.9 [24.8–34.6]	31.5 [27.5–35.2]	22.6 [17.6–28.0]
% of the reference VO_{2peak} *	92.9 [69.0–116.3]	91.4 [65.6–116.3]	99.5 [77.4–115.6]
Questionnaire data			
Time between CPET and questionnaires (days)	0 [−5–0]	0 [−6–0]	0 [−1–0]
FitMáx-VO_{2peak} (ml·kg^{-1}·min^{-1}) *	25.1 [18.6–38.3]	30.1 [19.1–39.5]	20.9 [15.8–27.4]
DASI-VO_{2peak} (ml·kg^{-1}·min^{-1}) *	31.2 [21.9–34.6] **	31.4 [22.2–34.6]**	25.4 [19.6–32.0]
VSAQ-VO_{2peak} (ml·kg^{-1}·min^{-1}) *	22.4 [15.1–36.9] **	29.0 [15.7–37.3]**	18.6 [14.2–22.7]**
MET questionnaire-VO_{2peak} (ml·kg^{-1}·min^{-1}) *	24.5 [14.0–28.0] **	24.5 [17.5–35.0]**	21.0 [14.0–24.5]
M-DASI-4Q *	3 [1–4]	3 [2–4]	2 [1–4]

Results are displayed as *n* (%) or median [IQR]

Missing information, number of subjects: FEV_1, 2; FVC, 2; DASI-VO_{2peak}, 120; VSAQ-VO_{2peak}, 39; MET questionnaire-VO_{2peak}, 50

* Significantly different between female and male subjects. In case of categorical values, a post hoc analysis with Bonferroni correction for multiple testing was performed

** Significantly different from CPET

[a] The prediction model for VO_{2peak} of the LowLands Fitness Registry was used as a reference value[45]

Abbreviations: ASA, American Society of Anesthesiologists physical status; BMI, body mass index; COPD, chronic obstructive pulmonary disease; CPET, cardiopulmonary exercise testing; DASI, duke activity status index; FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; GOLD, Global initiative for chronic Obstructive Lung Disease; HR_{peak}, heart rate at peak exercise; M-DSAI-4Q, modified 4-question duke activity status index; MET, metabolic equivalent of task; RER_{peak}, respiratory exchange ratio at peak exercise; VO_{2peak}, oxygen uptake at peak exercise; VSAQ, veterans-specific activity questionnaire; WR_{peak}, work rate at peak exercise

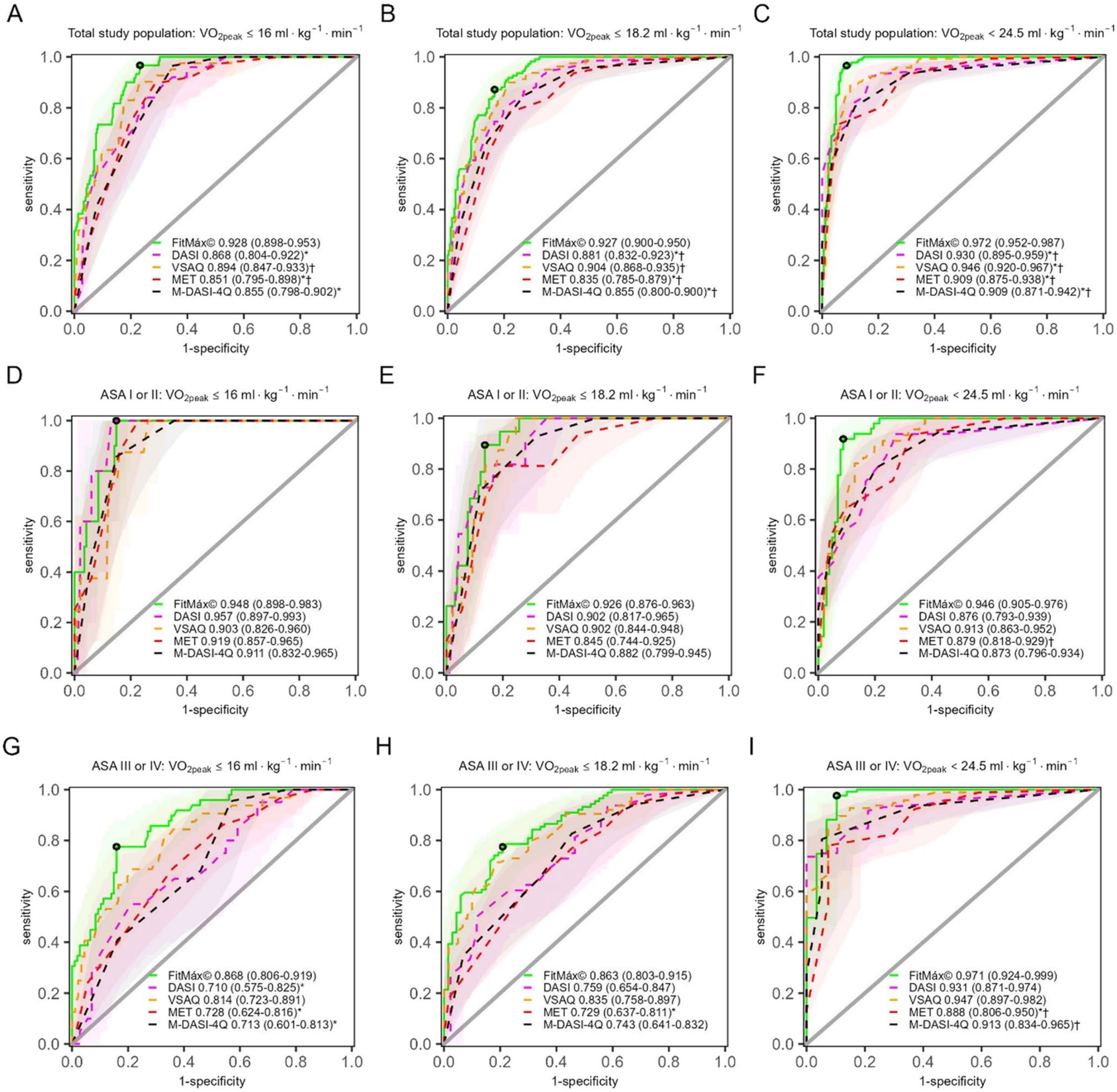

Fig. 1 ROC curves for detecting predefined VO_{2peak} thresholds with patient-reported questionnaires. **Notes:** AUC is reported with 95% CI. Graphs a-c: ROC curves to detect $VO_{2peak} \leq 16.0, \leq 18.2$ and ≤ 24.5 ml·kg^{-1}·min^{-1} in the total study population ($n = 361$). Graphs d-f: ROC curves to detect $VO_{2peak} \leq 16.0, \leq 18.2$ and ≤ 24.5 ml·kg^{-1}·min^{-1} in a subgroup with ASA I-II ($n = 151$). Graphs g-i: ROC curves to detect $VO_{2peak} \leq 16.0, \leq 18.2$ and ≤ 24.5 ml·kg^{-1}·min.$^{-1}$ in a subgroup with ASA III-IV ($n = 156$). ° Youden index of the FitMáx. * Significant different AUC compared to the FitMáx (complete case analysis). † Significant different Youden J-statistic compared to the FitMáx (complete case analysis)

(> 87%), and high negative predictive values (NPV > 90%) across all predefined values of the CPET-VO_{2peak} thresholds (Table 2) [36]. The FitMáx cut-off value corresponding to the VO_{2peak} threshold ≤ 24.5 ml·kg^{-1}·min^{-1}, with highest specificity and positive predictive value (PPV), was 26.1 ml·kg^{-1}·min^{-1}. Using this cut-off, the FitMáx achieved a sensitivity of 97%, specificity of 91%, PPV of 91%, and NPV of 97%. The corresponding cut-off value for VSAQ was 20.4 ml·kg^{-1}·min^{-1}, with a sensitivity of 89%, specificity of 90%, PPV of 88%, and NPV of 92%, which are slightly lower compared to the FitMáx. The analogous optimal cut-off value of the DASI was 29.2 ml·kg^{-1}·min^{-1}, and resulted in a sensitivity of 92%, specificity of 82%, PPV of 81%, and NPV of 93%, which are lower compared to both FitMáx

Table 2 Questionnaire cut-off values to detect a $VO_{2peak} \leq 16.0, \leq 18.2$ and ≤ 24.5 ml·kg^{-1}·min^{-1} based on Youden index in the total study population ($n = 361$)

Questionnaire and predefined VO_{2peak} threshold	Questionnaire Cut-off value	Sensitivity	Specificity	PPV	NPV
FitMáx					
≤ 16.0 ml·kg^{-1}·min^{-1}	20.6 ml·kg^{-1}·min^{-1}	0.97	0.77	0.45	0.99
≤ 18.2 ml·kg^{-1}·min^{-1}	21.3 ml·kg^{-1}·min^{-1}	0.87	0.83	0.69	0.94
≤ 24.5 ml·kg^{-1}·min^{-1}	26.1 ml·kg^{-1}·min^{-1}	0.97	0.91	0.91	0.97
DASI					
≤ 16.0 ml·kg^{-1}·min^{-1}	22.9 ml·kg^{-1}·min^{-1}	0.84	0.75	0.28	0.98
≤ 18.2 ml·kg^{-1}·min^{-1}	24.1 ml·kg^{-1}·min^{-1}	0.80	0.80	0.57	0.92
≤ 24.5 ml·kg^{-1}·min^{-1}	29.2 ml·kg^{-1}·min^{-1}	0.92	0.82	0.81	0.93
M-DASI-4Q					
≤ 16.0 ml·kg^{-1}·min^{-1}	2.0 [a]	0.97	0.65	0.26	0.99
≤ 18.2 ml·kg^{-1}·min^{-1}	2.0 [a]	0.85	0.73	0.52	0.93
≤ 24.5 ml·kg^{-1}·min^{-1}	2.0 [a]	0.81	0.88	0.84	0.85
VSAQ					
≤ 16.0 ml·kg^{-1}·min^{-1}	16.3 ml·kg^{-1}·min^{-1}	0.90	0.77	0.36	0.98
≤ 18.2 ml·kg^{-1}·min^{-1}	18.2 ml·kg^{-1}·min^{-1}	0.90	0.79	0.59	0.96
≤ 24.5 ml·kg^{-1}·min^{-1}	20.4 ml·kg^{-1}·min^{-1}	0.89	0.90	0.88	0.92
MET questionnaire					
≤ 16.0 ml·kg^{-1}·min^{-1}	17.5 ml·kg^{-1}·min^{-1}	0.89	0.71	0.30	0.98
≤ 18.2 ml·kg^{-1}·min^{-1}	17.5 ml·kg^{-1}·min^{-1}	0.78	0.78	0.54	0.91
≤ 24.5 ml·kg^{-1}·min^{-1}	17.5 ml·kg^{-1}·min^{-1}	0.74	0.94	0.90	0.82

[a] Predefined VO_{2peak} thresholds and Questionnaire cut-off values are expressed in ml·kg^{-1}·min^{-1} except for the M-DASI-4Q

Abbreviations: DASI, duke activity status index; M-DSAI-4Q, modified 4-question duke activity status index; MET, metabolic equivalent of task; NPV, negative predictive value; PPV, positive predictive value; VO_{2peak}, oxygen uptake at peak exercise; VSAQ, veterans-specific activity questionnaire

and VSAQ. These results were generally the same for the VSAQ and DASI in relation to the FitMáx for the two other VO_{2peak} thresholds (i.e., ≤ 18.2 and ≤ 24.5 ml·kg^{-1}·min^{-1}) (Table 2). For the MET-questionnaire and the M-DASI-4Q, the cut-off values of 17.5 ml·kg^{-1}·min^{-1} and 2.0 (on a scale 1–4) were found (due to the nature of these questionnaires), corresponding to all three predefined VO_{2peak} thresholds.

Table 3 presents FitMáx cut-off values corresponding to the predefined VO_{2peak} thresholds for predetermined sensitivity values. This enables clinicians to select the most suitable FitMáx-VO_{2peak} cut-off values in specific clinical situations.

The subgroup analysis based on ASA physical status classification consisted of 151 participants (104 men and 47 women) with ASA I-II, and a median (IQR) VO_{2peak} of 34.3 (27.1–42.7) ml•kg^{-1}•min^{-1} and 22.5 (18.7–27.0) ml•kg^{-1}•min^{-1}, for men and women respectively. The subgroup of ASA III-IV consisted of 156 participants (125 men and 31 women), with a median (IQR) VO_{2peak} of 17.8 (16.0–23.7) ml•kg^{-1}•min^{-1} and 16.2 (13.6–18.0) ml•kg^{-1}•min^{-1}, for men and women respectively.

AUC values of the questionnaires did not show a significant difference, when compared pairwise to the FitMáx in the subgroup with ASA I-II (Fig. 1d-f). In the subgroup with ASA III-IV, FitMáx showed a significantly higher AUC compared to the DASI, MET-questionnaire, and M-DASI-4Q for the VO_{2peak} threshold ≤ 16.0 ml•kg^{-1}•min^{-1}. The AUC for the FitMáx was not significantly different compared to the VSAQ (Fig. 1g). For the VO_{2peak} thresholds of ≤ 18.2 and ≤ 24.5 the AUC of the FitMáx showed only a significant difference when compared to the MET-questionnaire (Fig. 1h-i). As expected, the 95% CI are larger in the subgroup analyses compared to the total study population (Fig. 1).

For the subgroup with ASA III-IV, different optimal questionnaire cut-off values were found based on the Youden index, compared to the total study population (supplementary Table 1). For the FitMáx and DASI, the optimal cut-off values are closer to the predefined VO_{2peak} thresholds compared to the other questionnaires. For the VSAQ, the optimal cut-off values are lower compared to the predefined VO_{2peak} thresholds. The M-DASI-4Q and the MET-questionnaire showed the same cut-off value as in the total study population, except for the MET-questionnaire on the VO_{2peak} threshold of ≤ 16.0 ml•kg^{-1}•min^{-1}, for which a cut-off value of 14.0 ml•kg^{-1}•min^{-1} was found. Overall, in the subgroup with ASA III-IV, lower sensitivity values and NPV were found for all questionnaires to distinguish patients at risk. On the other hand, specificity and PPV were generally higher in the ASA III-IV subgroup, compared to the

Table 3 Cut-off values for the FitMáx to detect a $VO_{2peak} \leq 16.0$, ≤ 18.2 and ≤ 24.5 ml·kg^{-1}·min^{-1} based on sensitivity values in total study population ($n = 361$)

Sensitivity	$VO_{2peak} \leq 16.0$ ml·kg^{-1}·min^{-1}				$VO_{2peak} \leq 18.2$ ml·kg^{-1}·min^{-1}				$VO_{2peak} \leq 24.5$ ml·kg^{-1}·min^{-1}			
	FitMáx VO_{2peak} cut-off value (ml·kg^{-1}·min^{-1})	*Specificity*	*PPV*	*NPV*	*FitMáx VO_{2peak} cut-off value (ml·kg^{-1}·min^{-1})*	*Specificity*	*PPV*	*NPV*	*FitMáx VO_{2peak} cut-off value (ml·kg^{-1}·min^{-1})*	*Specificity*	*PPV*	*NPV*
0.05	10.3	1.00	1.00	0.84	10.8	1.00	1.00	0.71	11.5	1.00	1.00	0.52
0.10	10.8	1.00	1.00	0.85	12.1	1.00	1.00	0.72	13.1	1.00	1.00	0.54
0.15	11.5	1.00	1.00	0.86	12.9	1.00	1.00	0.73	14.7	1.00	1.00	0.55
0.20	12.1	1.00	1.00	0.86	13.4	1.00	1.00	0.74	15.6	0.99	0.97	0.56
0.25	12.7	1.00	1.00	0.87	15.0	0.99	0.91	0.76	16.0	0.99	0.96	0.58
0.30	12.9	1.00	1.00	0.88	15.6	0.99	0.92	0.77	16.4	0.99	0.96	0.60
0.35	14.3	0.99	0.88	0.88	15.8	0.98	0.91	0.78	17.0	0.99	0.97	0.61
0.40	15.1	0.97	0.73	0.89	16.2	0.97	0.87	0.79	17.8	0.99	0.97	0.63
0.45	15.6	0.96	0.69	0.90	16.4	0.97	0.88	0.81	18.5	0.98	0.95	0.65
0.50	15.8	0.95	0.68	0.91	17.0	0.96	0.86	0.82	18.9	0.98	0.96	0.67
0.55	16.3	0.93	0.62	0.91	17.5	0.96	0.86	0.83	19.2	0.98	0.96	0.70
0.60	16.4	0.93	0.63	0.92	18.6	0.92	0.76	0.84	19.7	0.97	0.96	0.72
0.65	16.8	0.92	0.63	0.93	18.7	0.92	0.77	0.86	20.1	0.97	0.95	0.74
0.70	17.1	0.92	0.64	0.94	19.1	0.91	0.77	0.88	21.0	0.96	0.94	0.77
0.75	18.6	0.86	0.53	0.95	19.4	0.90	0.77	0.89	21.7	0.95	0.94	0.80
0.80	18.7	0.86	0.53	0.96	20.4	0.85	0.71	0.92	22.3	0.95	0.94	0.83
0.85	19.1	0.83	0.50	0.97	21.2	0.83	0.69	0.93	22.9	0.95	0.94	0.87
0.90	19.5	0.82	0.50	0.98	22.3	0.79	0.66	0.95	23.8	0.94	0.94	0.91
0.95	20.4	0.77	0.45	0.99	24.3	0.72	0.60	0.98	25.3	0.92	0.92	0.96
1.00	22.3	0.70	0.40	1.00	27.6	0.67	0.57	1.00	30.0	0.85	0.87	1.00

Predefined VO_{2peak} thresholds and FitMáx cut-off values are expressed in ml·kg^{-1}·min^{-1}

Abbreviations: NPV, negative predictive value; PPV, positive predictive value; VO_{2peak}, oxygen uptake at peak exercise total study population. To reach a predetermined sensitivity of 90%, cut-off values for the FitMáx were 19.3, 22.2, and 23.4 ml·kg^{-1}·min^{-1} corresponding to VO_{2peak} thresholds of ≤ 16.0, ≤ 18.2, and ≤ 24.5 ml·kg^{-1}·min^{-1}, respectively.

Discussion

The current study investigated whether self-reported questionnaires could be used in the preoperative evaluation to identify high-risk surgical patients, using predefined clinically relevant VO_{2peak} thresholds corresponding to a higher risk for complications. Overall, all questionnaires demonstrated good ability to identify patients at high preoperative risk (with the VO_{2peak} thresholds as standard criterion) in the current study population [46]. However, differences in sensitivity, specificity, NPV, and PPV exist and should be taken into account in specific clinical situations. Not all questionnaires may be suitable in clinical practice, since their cut-off values may lead to many false positive outcomes, and patients may be referred for additional exercise testing or preventive measures unnecessarily. The FitMáx and VSAQ performed superiorly compared to the DASI, M-DASI-4Q, and MET questionnaire. In previous studies, DASI, VSAQ and FitMáx have shown to diverge in performance (correlation, bias and 95% limits of agreement) in predicting CPET-VO_{2peak} in several study populations [12, 23, 25, 28–31]. Struthers et al. reported poor discriminative ability of the DASI between patients with higher and lower preoperative risk, which may be caused by the relatively low mean VO_{2peak} of 18.7 ml·kg^{-1}·min^{-1} in their study population [29]. Wijeysundera et al. reported that subjective assessment with a MET categorization failed to detect low CRF in a study population with a mean VO_{2peak} of 19.2 ml·kg^{-1}·min^{-1} and a small range in VO_{2peak} [12]. The small range in VO_{2peak} is mostly due to a higher mean age and considerably less healthy participants in the study groups as opposed to the current study (median VO_{2peak} of 25.3 ml·kg^{-1}·min^{-1}) [12, 29]. Possibly, the presence of more fit subjects in a population enlarges the discriminative ability of questionnaires to detect subjects with lower CRF and concomitant higher preoperative risk for complications [2–4]. A lower discriminative ability of questionnaires to identify patients with low CRF, is also found in the current study in the subgroup with ASA III-IV.

In this study, AUC values to classify subjects based on VO_{2peak} thresholds were highest for the FitMáx for all three considered VO_{2peak} thresholds in the total study population and in the subgroups based on the ASA classification. However, the AUC values of the FitMáx were not always *significantly* higher compared to the other questionnaires. Also, optimal questionnaire cut-off values based on the Youden index were presented. However, the Youden index does not distinguish between the importance of sensitivity versus specificity in clinical situations [36]. Therefore, an additional analysis based on predetermined sensitivity values was performed [38]. To reach a sensitivity of 90%, the FitMáx cut-off values of 19.5, 22.3, and 23.8 ml·kg^{-1}·min^{-1} were found in correspondence with VO_{2peak} thresholds of ≤ 16.0, ≤ 18.2 and ≤ 24.5 ml·kg^{-1}·min^{-1}, respectively. Depending on the VO_{2peak} threshold that is used in clinical practice, preoperative CPET may be unnecessary for patients with VO_{2peak} estimated from the FitMáx questionnaire above these cut-off values, unless an underlying cardiac or pulmonary exercise limitation is suspected.

Subgroup analyses based on ASA classification showed comparable AUC values for ASA I-II, and lower AUC values with wider 95%-CI for ASA III-IV to detect the predefined VO_{2peak} thresholds, compared to the total study population [46]. As described earlier, this leads to the assumption that the discriminative ability of the questionnaires is lower in a population with lower mean CRF [29]. Based on the sensitivity analysis in the ASA III-IV subgroup, the same FitMáx cut-off value was found to detect VO_{2peak} ≤ 18.2 ml·kg^{-1}·min^{-1}, and slightly different cut-off values were found to detect VO_{2peak} ≤ 16.0 and ≤ 24.5 ml·kg^{-1}·min^{-1}, compared to the total study population. However, specificity and PPV are lower compared to the total study population, indicating more false-positive and less true-positives cases. This means being classified as less fit because of a low FitMáx-VO_{2peak}, in the ASA III-IV subgroup, does not necessarily mean a person is truly less fit [37].

Remarkably, the optimal cut-off values of the MET-questionnaire based on the Youden index are the same for all predefined VO_{2peak} thresholds, except for the cut-off value corresponding to the threshold VO_{2peak} ≤ 16.0 ml·kg^{-1}·min^{-1} in the ASA III-IV subgroup [36]. The cut-off value of VO_{2peak} ≤ 17.5 ml·kg^{-1}·min^{-1} corresponds to the METs cut-off value of 7, as suggested by the American Heart Association [17]. Based on the Youden index, the M-DASI-4Q only showed one possible cut-off value of 2.0 (on a scale 1–4). This is probably due to the sole purpose of the questionnaire to provide a preoperative risk cut-off point [36]. Based on the current study results, it is recommended to use the FitMáx or VSAQ in preoperative screening to distinguish patients at a higher risk based on predefined VO_{2peak} thresholds. However, these self-reported questionnaires do not take patients' comorbidities into account. It is therefore advised to combine subjective assessment by anaesthesiologists/physicians with the questionnaire-predicted CRF to enable informed decision-making [11]. Patients with a poor result (predicted VO_{2peak} ≤ 21.3 and ≤ 18.2 ml·kg^{-1}·min^{-1} respectively for FitMáx and VSAQ) could benefit from further preoperative (cardiopulmonary) exercise testing and preventive interventions (e.g., prehabilitation).

Study limitations and implications for future research

The main limitation of the current study is that data was used from a cohort of participants recruited in a general clinical setting instead of a specific preoperative setting. Although ASA III was the most common physical status classification in the current study population, indicating patients with severe systemic diseases[11], one might question to which extent results can be extrapolated to a preoperative population, since postoperative follow-up was not possible. Future studies should focus on the direct relationship of questionnaire-estimated preoperative VO_{2peak} with peri- and postoperative outcomes (e.g. length of hospitalization, complications, mortality) in a surgical study population to validate the role of questionnaires in preoperative risk screening and/or assessment. Moreover, the current study population consisted of mainly male subjects (77%). A recent study reported a sex-specific difference in VO_{2peak} thresholds for predicting postoperative complications [47]. To enhance the interpretability for female subjects, more data should be collected to evaluate the accuracy of detecting sex-specific VO_{2peak} thresholds. A limitation of the questionnaires is the inclusion of activities that are not globally practiced (e.g. cycling, basketball, skiing) [22, 27]. However, in the previous validation study, the predicted VO_{2peak} based on the FitMáx without cycling showed similar results compared to the FitMáx with cycling [22].

Conclusion

The FitMáx, DASI, M-DASI-4Q, VSAQ, and MET questionnaire all showed high AUC values in a general Dutch population to preoperatively identify patients at high risk, in which the FitMáx and VSAQ performed superiorly. Both could be used as self-reported questionnaires to identify high-risk surgical patients who might benefit from formal preoperative (cardiopulmonary) exercise testing and preoperative interventions to reduce surgical risks.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00423-024-03560-0.

Acknowledgements We would like to thank S. Van Keeken and C.J.L. Molenaar for their assistance with the study.

Author's contributions R.M., D.C., M.H. and G.S. contributed to the study conception and design. R.M. and M.H. performed material preparation and data collection. All authors contributes to the analysis and interpretation of the data. R.M. wrote the first draft of the manuscript. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the by the Stichting Opleidingen in de Sportgezondheidszorg and by the National Foundation Against Cancer.

Data availability More information on the data gathered and the analysis used is available upon reasonable request.

Declarations

Ethics approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Ethical approval for this study (reference number N18.051) was provided by the authorised Medical Research Ethics Committee of the MMC, Veldhoven, the Netherlands on 30 April 2018. Informed consent for participation was obtained from all individual participants included in the study.

Competing interests The authors declare no competing interests.

References

1. Ross R, Blair SN, Arena R et al (2016) Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. Circulation 134(24):e653–e699
2. Heldens A, Bongers BC, Lenssen AF, Stassen LPS, Buhre WF, van Meeteren NLU (2017) The association between performance parameters of physical fitness and postoperative outcomes in patients undergoing colorectal surgery: An evaluation of care data. Eur J Surg Oncol 43(11):2084–2092
3. Snowden CP, Prentis J, Jacques B et al (2013) Cardiorespiratory fitness predicts mortality and hospital length of stay after major elective surgery in older people. Ann Surg 257(6):999–1004
4. Steffens D, Ismail H, Denehy L et al (2021) Preoperative Cardiopulmonary Exercise Test Associated with Postoperative Outcomes in Patients Undergoing Cancer Surgery: A Systematic Review and Meta-Analyses. Ann Surg Oncol 28(12):7120–7146
5. Thomas G, Tahir MR, Bongers BC, Kallen VL, Slooter GD, van Meeteren NL (2019) Prehabilitation before major intra-abdominal cancer surgery: A systematic review of randomised controlled trials. Eur J Anaesthesiol 36(12):933–945
6. Berkel AEM, Bongers BC, Kotte H et al (2022) Effects of Community-based Exercise Prehabilitation for Patients Scheduled for Colorectal Surgery With High Risk for Postoperative Complications: Results of a Randomized Clinical Trial. Ann Surg 275(2):e299–e306
7. Molenaar CJ, van Rooijen SJ, Fokkenrood HJ, Roumen RM, Janssen L, Slooter GD (2022) Prehabilitation versus no prehabilitation to improve functional capacity, reduce postoperative complications and improve quality of life in colorectal cancer surgery. Cochrane Database Syst Rev 5(5):Cd013259

8. Tew GA, Ayyash R, Durrand J, Danjoux GR (2018) Clinical guideline and recommendations on pre-operative exercise training in patients awaiting major non-cardiac surgery. Anaesthesia 73(6):750–768
9. Molenaar CJL, Minnella EM, Coca-Martinez M et al (2023) Effect of Multimodal Prehabilitation on Reducing Postoperative Complications and Enhancing Functional Capacity Following Colorectal Cancer Surgery: The PREHAB Randomized Clinical Trial. JAMA Surg 158:572–581
10. Saklad M (1941) Grading Of Patients For Surgical Procedures. Anesthesiology 2(3):281–284
11. Mayhew D, Mendonca V, Murthy BVS (2019) A review of ASA physical status - historical perspectives and modern developments. Anaesthesia 74(3):373–379
12. Wijeysundera DN, Pearse RM, Shulman MA et al (2018) Assessment of functional capacity before major non-cardiac surgery: an international, prospective cohort study. Lancet 391(10140):2631–2640
13. Levett DZ, Grocott MP (2015) Cardiopulmonary exercise testing for risk prediction in major abdominal surgery. Anesthesiol Clin 33(1):1–16
14. Stubbs DJ, Grimes LA, Ercole A (2020) Performance of cardiopulmonary exercise testing for the prediction of post-operative complications in non cardiopulmonary surgery: A systematic review. PLoS One 15(2):e0226480
15. Older PO, Levett DZH (2017) Cardiopulmonary Exercise Testing and Surgery. Ann Am Thorac Soc 14(Supplement_1):S74-s83
16. West MA, Asher R, Browning M et al (2016) Validation of preoperative cardiopulmonary exercise testing-derived variables to predict in-hospital morbidity after major colorectal surgery. Br J Surg 103(6):744–752
17. Fleisher LA, Fleischmann KE, Auerbach AD et al (2014) 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 130(24):e278-333
18. Shah SJ, Rehman A, Shaukat MHS, Awais M (2017) Cost-effectiveness of exercise stress testing performed as part of executive health examinations. Ir J Med Sci 186(2):281–284
19. Reeves T, Bates S, Sharp T et al (2018) Cardiopulmonary exercise testing (CPET) in the United Kingdom-a national survey of the structure, conduct, interpretation and funding. Perioper Med (Lond) 7:2
20. Keteyian SJ, Brawner CA, Ehrman JK, Ivanhoe R, Boehmer JP, Abraham WT (2010) Reproducibility of peak oxygen uptake and other cardiopulmonary exercise parameters: implications for clinical trials and clinical practice. Chest 138(4):950–955
21. Melon CC, Eshtiaghi P, Luksun WJ, Wijeysundera DN (2014) Validated questionnaire vs physicians' judgment to estimate preoperative exercise capacity. JAMA Intern Med 174(9):1507–1508
22. Meijer R, van Hooff M, Papen-Botterhuis NE et al (2022) Estimating VO(2peak) in 18–90 Year-Old Adults: Development and Validation of the FitMáx©-Questionnaire. Int J Gen Med 15:3727–3737
23. Hlatky MA, Boineau RE, Higginbotham MB et al (1989) A brief self-administered questionnaire to determine functional capacity (the Duke Activity Status Index). Am J Cardiol 64(10):651–654
24. Riedel B, Li MH, Lee CHA et al (2021) A simplified (modified) Duke Activity Status Index (M-DASI) to characterise functional capacity: a secondary analysis of the Measurement of Exercise Tolerance before Surgery (METS) study. Br J Anaesth 126(1):181–190
25. Myers J, Do D, Herbert W, Ribisl P, Froelicher VF (1994) A nomogram to predict exercise capacity from a specific activity questionnaire and clinical data. Am J Cardiol 73(8):591–596
26. Ainsworth BE, Haskell WL, Herrmann SD et al (2011) 2011 Compendium of Physical Activities: a second update of codes and MET values. Med Sci Sports Exerc 43(8):1575–1581
27. Kojima S, Wang DH, Tokumori K et al (2006) Practicality of Veterans Specific Activity Questionnaire in evaluation of exercise capacity of community-dwelling Japanese elderly. Environ Health Prev Med 11(6):313–320
28. Li MH, Bolshinsky V, Ismail H, Ho KM, Heriot A, Riedel B (2018) Comparison of Duke Activity Status Index with cardiopulmonary exercise testing in cancer patients. J Anesth 32(4):576–584
29. Struthers R, Erasmus P, Holmes K, Warman P, Collingwood A, Sneyd JR (2008) Assessing fitness for surgery: a comparison of questionnaire, incremental shuttle walk, and cardiopulmonary exercise testing in general surgical patients. Br J Anaesth 101(6):774–780
30. van Leunen MMCJ, van Hooff M, Meijer R et al (2023) Evaluation of two self-reported questionnaires for assessment of peak exercise capacity in patients with chronic heart failure. Eur J Prev Cardiol 31:e13–e15
31. Weemaes ATR, Meijer R, Beelen M et al (2023) Monitoring aerobic capacity in cancer survivors using self-reported questionnaires: criterion validity and responsiveness. J Patient-Rep Outcomes 7(1):73
32. Meijer R, Schep G, Regis M, Papen-Botterhuis NE, Savelberg HHCM, van Hooff M (2024) Test–retest reliability of the FitMáx©-questionnaire in a clinical and healthy population. J Patient-Rep Outcomes 8(1):3
33. Levett DZH, Jack S, Swart M et al (2018) Perioperative cardiopulmonary exercise testing (CPET): consensus clinical guidelines on indications, organization, conduct, and physiological interpretation. Br J Anaesth 120(3):484–500
34. Kim HY (2013) Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restor Dent Endod 38(1):52–54
35. Hanley JA, McNeil BJ (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148(3):839–843
36. Youden WJ (1950) Index for rating diagnostic tests. Cancer 3(1):32–35
37. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R (2008) Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol 56(1):45–50
38. Pepe M, Longton G, Janes H (2009) Estimation and Comparison of Receiver Operating Characteristic Curves. Stata J 9(1):1
39. Trevethan R (2017) Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. Front Public Health 5:307
40. Bantis LE, Nakas CT, Reiser B (2021) Statistical inference for the difference between two maximized Youden indices obtained from correlated biomarkers. Biom J 63(6):1241–1253
41. Benjamini Y, Hochberg Y (2018) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc: Ser B (Methodol) 57(1):289–300
42. Wickham H (2022) Tools for Splitting, Applying and Combining Data 1.8.8 ed. https://doi.org/10.32614/CRAN.package.plyr
43. Robin X, Turck N, Hainard A et al (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12:77
44. Team RC (2020) R: A language and environment for statistical computing. 4.0 ed. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
45. van der Steeg GE, Takken T (2021) Reference values for maximum oxygen uptake relative to body mass in Dutch/Flemish subjects aged 6–65 years: the LowLands Fitness Registry. Eur J Appl Physiol 121(4):1189–1196

46. Nahm FS (2022) Receiver operating characteristic curve: overview and practical use for clinicians. Korean J Anesthesiol 75(1):25–36
47. Alfitian J, Riedel B, Ismail H et al (2023) 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. EClinicalMedicine 64:102223