Peer review file

Article information: http://dx.doi.org/10.21037/jtd-20-1528

Reviewer A

The authors performed a retrospective, observational study on CPET for pre-operative risk stratification among patients with NSCLC. This is an important paper and I enjoyed reading it. The paper's strengths are its sample size, adherence to STROBE guidelines, and comprehensive data collection. Weaknesses are the failure to cite the reference equations, over-simplification of exercise physiology (common in papers on CPET), and recommendations based on assumptions that were not proven by their data. I would be willing to read this paper again, if the below issues can satisfactorily be addressed.

Comment 1:

Introduction: The authors should check the primary data (study included in reference 8, the meta-analysis) to ensure that VO2max obtained in studies that predict post-surgical outcomes was in fact VO2max, and not peak. There's no reason to think that the patients enrolled in most of those studies were anymore capable of hitting a VO2max than were the patients they studied for this paper.

Reply 1:

Thank you for this relevant comment. As suggested, we checked the 14 studies (listed below) included in the meta-analysis of Benzo published in 2007.

Of the 14 studies included in the meta-analysis, seven of them only refer to the concept of "VO2max". Among these, only one of them (Bolliger 1995) clearly mentions the maximum reached in the great majority of patients. Five mention "VO2max", but defining it as the maximum value reached during the CPET, which for many could be assimilated to VO2peak. Finally, two clearly express the use of VO2peak in their protocol. In total, the ambiguity between the predictive value of VO2peak or VO2max remains present.

The review of all these data which led to the generation of the cut-offs of 10 to 20 ml/kg/min of VO2 also leads us to consider that the patients referred for preoperative physiological evaluation are no longer the same as in the 80s to the beginning of the 2000s (references to average VO2max and the strong increment per minute in some studies shown in the table below). Note that several of these studies took place in military hospitals with patients with a sporting past, and despite their dysfunction, probably an ability to mentally tolerate physical effort and not to stop prematurely.

The patients addressed today, thanks to the advances in surgery and anaesthesia are becoming more severe and poly-pathological (see additional data on co-morbidities and medical treatments). These are currently particularly frail for the most part and for whom the practice of a present or past physical activity is rare.

Regarding the evolution of the type of patients referred for CPET before thoracic surgery, the question of the method of evaluation and its limits in some of them seems topical to us. This study remains limited in terms of level of evidence due to its retrospective aspect and the lack of correlation on complications (as mentioned by reviewer n°2). This would indeed merit future complementary research.

Study	N=	Incremental protocol	VO2peak or VO2max ?	Mean VO2max/peak
Bechard (1987)	50 "Hunter Holmes McGuire Veterans Administration Medical Center"	"multistage incremental programmed protocol (…) of one-minute 12.5-W (…) until the patient was unable to continue because of exhaustion or dyspnea."	"Maximal O2 consumption (MVo2) based on body weight was expressed in milligrams per kilogram per minute and was defined as the highest Vo2 obtained during the exercise study"	Vo2max, ml/kg/min 17.01 ± 0.77 9.95 ± 1.52, p<0.001 without complication vs with complication respectively
Bolliger (1995)	80	"2 min warm-up period at 20 W, a ramp protocol with a 20 W/min" (...) "The exercise test was stopped when the patients were exhausted, a plateau in v02uptake appeared, or at any signs (electrocardiogram ECG) or symptoms of myocardial ischemia, including a fall in blood pressure."	"most patients reached a plateau in their oxygen consumption curve or developed a metabolic acidosis at peak exercise"	Vo2max, ml/kg/min 19.7 ± 5 vs 14.7 ± 3, p=0.0002 without complication vs with complication respectively
Boysen (1990)	17 « Gainesville (FL) Veterans Administration Medical Center »	« multistage incremental treadmill testing to maximum performance. Treadmill speed began at 2 mph with zero grade. At l-minute intervals, the speed was increased by 1 mph. The test was terminated at 12 minutes, or sooner if the subject was unable to continue." (…) "the study protocol was incremental	only evokes the notion of VO2max	VO2max (mL/kg/min) 20.63 ± 5.53 vs 17.00 ± 4.30 without complication vs with complication respectively

		but was designed to reach maximal workloads and to maintain a steady state for at least 1 minute."		
Brunelli (2002)	160	No CPET => estimate of VO2 from work (in watts), itself already estimated on the basis of an equation allowing to estimate work in watts according to the number of steps climbed?	only evokes the notion of VO2max (no measure of VO2 during exercise, only estimation)	Vo2max (mL/min/kg) 23.61 (4.0) ± 25.85 (4.4), p=0.03 with complication vs without complication respectively
Brutsche	125	« All patients underwent symptom-limited CPETs (…)2-min warm-up period at 20 W, a ramp protocol with 20-W.min-1. (…)The exercise test was stopped when the patients were exhausted, or at any electrocardiographic signs or clinical symptoms of myocardial ischaemia, including a fall in systolic blood pressure."	only evokes the notion of VO2max	Vo2max (mL/min/kg) 22.1±5.4 vs 17.9±5.4, p=0.0003 without complication vs with complication respectively
Epstein (1993)	42	"Symptom-limited exercise tests were performed using a ramp protocol on a bicycle ergometer. with the work rate determined using the method of Wasserman et al. Peak Vo2, and heart rate (HR) were averaged during the last 30 s of exercise ."	VO2peak averaged during the last 30 s of exercise	VO2peak (ml/min/kg) 16.4± 1.4 16.7±0.8, p= NS with complication vs without complication respectively
Larsen (1997)	97	« A multistage ramp, 10–15 W·min-1 maximal exercise	only evokes the notion of VO2max	Vo2max (mL/min)

		protocol (…) All patients were monitored with a 12-lead electrocardiogram (ECG) during exercise tests. Hard copies were produced at rest, every minute during work, at maximum exercise, and when arrhythmias or other ECG changes occurred. Anaerobic threshold (AT) corresponding to the maximum V 'O2 (VO2max; in mL·min-1) was estimated by the V-slope method »		1376 ± 412 for all patients Body weight (kg) 70.2 ± 12.2 So a mean VO2max around 19 ml/min/kg
Markos (1989)	53	« Work load (W) was increased each 1 minute by 100 kilopondmeters per minute (kpm-rnin'). (*around 16,3 watts/min*) (…) The patient was asked to exercise to maximal capacity, but the test was stopped if electrocardiographic abnormalities developed"	only evokes the notion of VO2max	Vo2max (mL/min/kg) Mean calculated only for patients with follow-up 22.1 ± 2.9 for pre-lobectomy n=4 16.6 ± 8.7 for pre-pneumonectomy
Morice (1992)	8	« The exercise protocol consisted of lo-W ramped increase in work every minute until the patient was unable to continue because of severe dyspnea or exhaustion. »	"peak oxygen uptake"	Vo2max (mL/min/kg) 16.7 ± 1.9 Patients with a VO2peak <15 ml/kg/min were excluded from the study and referred for appropriate nonsurgical therapy.
Smith (1984)	22 Shreveport Veterans Administration Hospital	The exercise protocol consisted of a 1-min 10-W incremental exercise	only evokes the notion of VO2max	Vo2max (mL/min/kg)

		test performed on a cycle ergometer (…) Maximal oxygen uptake (Vo2max) was defined as the highest Vo2 obtained during the exercise test.	but possibly VO2peak	22.4 ± 1.4 vs 14.9 ± 0.9, p<0.001 without complication vs with complication respectively
Torchio	56	Only abstract no information	Speak about VO2peak	Not available
Villani	150	Only abstract no information	only evokes the notion of VO2max	Not available
Wang 1999	40	« The patient's MVO2 was determined by a multi-stage incremental tes (…)The duration of exercise was determined by the physician administering the exercise study, and exercise took place at a standardized work rate (5, 10, 15, or 20 W/min) based on the patient's mass, height, age, and FEV1, as described by Wasserman and colleagues (…) The MVO2 was defined as the highest oxygen consumption achieved during the exercise test."	only evokes the notion of VO2max but possibly VO2peak	Vo2max (mL/min/kg) 17.8 ± 0.9 vs 16.3 ± 1.2, p=0.27 without complication vs with complication respectively no correlation between VO2 and complication
Wang 2000	57	« Maximal exercise capacity was determined with a stepped incremental exercise test protocol starting at a workload of 15 W, which was increased every minute by 15 W (…)he exercise test was discontinued when the patient reached 90% of the maximal predicted HR or felt fatigued and unable to continue, if an	only evokes the notion of VO2max but possibly VO2peak	Vo2max (mL/min/kg) 15.0 ± 2.4 vs 19.2 ± 4.3, p<0.001 with complication vs without complication respectively

	abnormal ECG developed, or if the SaO2 fell below 85%. The maximal oxygen uptake (VO2max) attained was taken to be the highest O2 consumption at the highest workload, just before the exercise test was discontinued.”

Comment 2:

Methods: Patients and Design

Were patients included consecutively?

Reply 2: Actually, although it is a retrospective study, it concerns patients screened in chronological order in the database of the physiology department of the Rouen university hospital centre and having carried out their CPET consecutively. This is now mentioned.

Changes in the text:

- **line 110:** “Patients with a diagnosis or suspicion of NSCLC and impaired pulmonary function or comorbidities, **consecutively** referred between January 2014 and July 2019,”

Comment 3:

Markers used to designate exercise test as normal:

Which reference sets were used for predicted max work rate and VO2?

Reply 3: We completely agree with this comment, this reference is important in terms of interpreting the data according to the reference equation. Thank you for raising this oversight. The predicted absolute VO2max, was derived from the equations reported by Wasserman et al.(13).

Changes in the text:

- **line 147:** “The predicted maximal work rate (Wmax) and V̇O2max was derived from the equations reported by Wasserman et al. (13).”.

Comment 4:

Results: Peak exercise values and maximality criteria:

Again – which reference set was used to establish predicted VO2?

Reply 4: This has been corrected in accordance with the previous comment.

Comment 5:

How was the diagnosis of peripheral muscle deconditioning made? Lines 224-228 on page 10 of the discussion seem to say this defined as difficulty pedalling and low AT. First, this definition should be stated earlier in the methods or when results are reported so the reader knows how its being defined. Second, we need to know how AT was defined. V-slope method? Ventilatory equivalents?

Reply 5:

First, the contribution of peripheral deconditioning has been repeatedly highlighted in the CPET conclusions by the physiologist on the basis of an early ventilatory threshold (<40% of theoretical VO2max) associated with poor performance achieved in terms of power reported to body-weight, perceived peripheral fatigue, +/- amyotrophy...

On a scientific level, we are in fact aware that this diagnosis is not based on additional examinations of finer and more invasive measurements. However, this is impossible within the framework of current clinical practice.

The reasons for exercise limitation are often multiple and interwoven within the frail patients included in this study. Given the presence of the factors previously mentioned, as well as the clinical examination preceding the CPET (interrogation highlighting a significant sedentary lifestyle, observation of muscle loss, etc.), the contribution of peripheral deconditioning seems highly probable.

- ⇨ Based on your recommendations, we have added additional elements in the result section

Second, concerning the de definition of the AT, it was manually determined as the average of the results obtained using the first break in VE, V-slope and ventilatory equivalents methods

Changes in the text:

- **Line 148:** "Ventilatory threshold (VT) was manually determined as the average of the results obtained using the first break in VE, V-slope and ventilatory equivalents methods (7)."
- **Lines 201 to 209:** "This was followed by peripheral muscle deconditioning, diagnosed on the basis of the conjunction of different factors such as an early ventilatory threshold, peripheral fatigue, poor performance in terms of power reached, muscle wasting, sedentary behaviour. Ventilatory threshold (VT) could not be measured for 12 patients and the ratio V̇O2 at VT / V̇O2max predicted was below 40% for 49 patients. This reflects a limitation in O2 supply to the tissues, compatible in part with deconditioning. Moreover, the median power/body-weight ratio was 1.1 watts/kilogram at peak exercise, which reflects poor performance in terms of power developed. Muscle deconditioning was rarely an isolated cause of limitation because it stemmed from health history."

Comment 6:

Was ventilatory limitation simply a produced of mechanical limitations (VE/MVV), or did gas-exchange abnormalities also make up some of the respiratory limitations seen?

Reply 6: The ventilatory limitation observed was mainly mechanical since only 37 (18.2%) patients presented a desaturation ≥ 4% during the CPET.

The patients who have desaturated are almost part of the patients who have entered in or used up their ventilatory reserve (34/37), the other presented diffusion disorder (DLCO < 50%).

Changes in the text:

- **Table 3:** addition of the number of patients with desaturation
- **Line 198:** "One hundred and four patients presented mechanical ventilatory limitation limitations with exhaustion of their VR. In addition, 37 patients presented gas-exchange impairment with desaturation ≥ 4% during the CPET. Most of them (34/37) were associated with an exhaustion of the VR, or deficit of diffusion observed on the PFT (3/37)."

Comment 7:

Discussion: The discussion of CV co-morbidities that begins on page 10, line 229 has numerous problems. First, authors have not defined when they would consider CV co-morbidities the cause of exercise limitation. Second, they infer that low heart-rate implies limitation due to CV co-morbidity, but this is often not the case. It often simply means poor effort. Low AT is part of a CV-disease pattern, and this goes unmentioned. Also, they mention patients on B-blockers. B-blocker use is a big confounder, and the presence of B-blocker use in this population is not surprising, it should make its appearance long before the discussion section. An expanded table 1 with medications and co-morbidities would go a long way toward solving this problem.

Reply 7: Thank you for raising this point, which we indeed have to agree, has been treated too lightly. We focused on the ventilatory and peripheral elements insofar as the cardiovascular limitations were primarily directed towards a cardiac rehabilitation centre, and were therefore limited in the acquisition of data.

On your initiative, we were able to review the cardiovascular limitations in more detail. These are now better defined and integrated into the results.

Likewise, a supplementary Table 1 with medications and comorbidities has been integrated as suggested.

Changes in the text:

- **Lines 210 to 219:** "Finally, cardiovascular function contributed to exercise limitation. Forty-two patients had peripheral arterial disease and 25 diagnosed heart disease (10 overlaps), which impacted their exercise performance. Although 35 patients were on beta-blocker, only 9 cases of chronotropic incompetence were highlighted as the main limiting factor. Fifteen patients presented an alteration of the ECG (sus/sub-ST, repolarization disorder, ventricular extrasystole) of which 7 required the early cessation of the effort, the others having occurred at near-maximum effort. Two cases of arterial hypertension (≥ 250 / 120 mmHg) necessitated to stop the CPET, one prematurely, and

one at near-maximum effort. A pathological kinetics of the O2 pulse were noted in 4 patients, suggesting an unknown onset heart failure."

Comment 8:

Strategies to reduce the risk of under-estimation: This was not the focus of their study. These can be mentioned in the discussion, but they should receive 1-2 sentences at the most. This section needs to be shortened.

Reply 8: Indeed, this part was not the focus of our study, however it seemed relevant for us to try to develop a clinical and practical message following the demonstration of the risk of underestimation of the VO2peak. This part can be expected or appreciated by some readers, as suggested by reviewer n°2. We take the decision not to shorten it since this is compatible with the number of words authorized by the publisher for this type of article.

Comment 9:

Conclusion:

The last few sentences of the conclusion need to be softened. They can state that their data suggest that some of the mitigation strategies they recommend may be useful. However, each of these strategies would require validation, and again, they didn't study any of them for this paper.

Reply 9: We take note of this comment. It is important to specify that these strategies remain suggestions on the basis of the literature and questioning in our practice. These are mere assumptions and deserve validation.

Changes in the text:

- **Line 392:** "These strategies must however be validated in order to help to ensure that appropriate surgical and therapeutic decisions are made for patients with NSCLC."

Comment 10:

Grammar

Page 5, line 101: "does" should be "do"

Page 5, line 103-104: "was" should be "were"

Page 6, line 141: "where" should be "were"

Page 8, line 172: "were" should be "was"

Page 9, line 202: The word "previous" here distorts this whole sentence. Do they mean "different" instead of previous? All patients only had one CPET, right? This is the first mention of "previous" studies. The table referred to seems to be a comparison across different ramps.

Reply 10: We thank you for these remarks and have made the changes indicated. We confirm that each patient has of course performed only one CPET. This paragraph being rightly mentioned as abrupt by reviewer 2, it has been completely revised.

Reviewer B

The introduction was well-written. It describes the scale of the issue to build up to the authors' hypothesis. The objectives were clearly stated which are to describe peak exercise variables and achievement of criteria defining peak exercise and to analyse the duration of incremental ramp and load increments. It is an interesting area surrounding CPET assessment, and not studied very much.

Comment 1:

The cohort was defined as patients with confirmed or suspected non-small cell lung cancer referred for surgical risk assessment using CPET, between January 2014 to July 2019. This is a study of relatively recent times (within the last 5 years). However, the authors did not clearly specify why this cohort was referred for risk assessment using CPET. Are they higher risk as defined by age, co-morbidity or impaired lung function alone? Is CPET routinely used as risk assessment in the authors' institute? The CPET protocol was very clearly defined, and only those who had undergone CPET using cycle ergometer were included. This was to standardise the results.

Reply 1: We thank you for this remark. As you suggest the reason for prescribing the preoperative CPET should be mentioned. These are patients considered fragile, either because of impaired respiratory function, or because of their comorbidities.

At the suggestion of reviewer n° 1, we inserted a supplementary Table 1 with comorbidities and medications. We hope this will also complement your request for information.

Changes in the text:

- **Line 110: "**Patients with a diagnosis or suspicion of NSCLC and impaired pulmonary function or comorbidities, consecutively referred between January 2014 and July 2019, for CPET to determine surgical risks at the Rouen University Hospital respiratory and exercise physiology department, were retrospectively included."
- **Table:** See supplementary Table 1

Comment 2:

The main disadvantage of this study is that there is no correlation of the authors' findings with clinical outcomes. How many of the patients who did not demonstrate maximal effort, and therefore had an overestimation of surgical risk and was declined for surgery? In those who ultimately undergone resection, what was their clinical outcomes in terms of 30-day mortality and post-operative complications, including complications leading to ITU admission and

invasive ventilation? Is the fitness for surgery made solely based on CPET results? Who ultimately make the decision regarding the patient's surgical risk?

Reply 2: We fully agree with this comment. The interest of the patient journey is very important. Whether it is surgery or even another therapy, it would be relevant to assess whether the physical capacity, as well as its reliability, could influence the therapeutic course.

Although this is a crucial element, we are unable to answer it at this time in view of the data in our possession. It is certainly a job that we should be looking for in the future.

Comment 3:

The cohort is a reasonable size in studies reporting on the use of CPET in lung resection. Figure 1 was a repetition of the first paragraph in the result section.

Reply 3: Figure 1 shows the same elements as the text. It seemed interesting for the reader to have a quick visual feedback. This may seem redundant, as you suggest. We leave the choice to the editor to indicate to us the maintenance or the withdrawal of this figure.

Comment 4:

It may be reasonable and useful to include post-operative predicted FEV1 and TLCO in Table 1. It would give the readers a better idea of the risk profile in terms of respiratory reserve.

Reply 4: We extracted the FEV1ppo and DLCOppo as suggested.

Changes in the text:

- **Line 132:** "Predictive postoperative (PPO) forced expiratory volume in 1 second (FEV1) and diffusing capacity of the lung for carbon monoxide (DLCO) were calculated according to the equations recommended by the ERS/ESTS guidelines (5)."
- **Table 1:** addition of FEV1ppo and DLCOppo

Comment 5:

Another minor remark I have is that it might flow better if the variables in Figure 2 is presented in the same order it was described in the text when the criteria were defined.

Reply 5: We have made the change requested in figure 2 and kept the order of the text (identical to that of the 2003 ACCP recommendations).

Changes in the text:

- **Figure 2:** modified

Comment 6:

Otherwise, there is good description and correlation of the markers/criteria achieved. The text and graphs illustrated the points well. This study has a significant finding, which is majority of

patients did not achieve real VO2 max. The cohort is not particularly old as median age was 66. CPET protocol used here is cycle-ergometer, which is a morefriendly technique as it is manageable by more patients, especially in those with arthritis or conditions limiting walking. Yet, despite the lack of obvious reasons in this cohort, it was found that about half of the cohort did not reach maximal effort.

Reply 6: Indeed, the majority of patients do not reach a "real" VO2max on the criterion of reaching a plateau or a VO2peak comparable to the theoretical maximum value. However, we show that some patients have a VO2peak associated with several maximality factors that can attest to a real sustained physiological stress, and for this we retain the need to present at least 2 associated maximality factors. Finally, around 30% of patients remain who do not seem to develop a really sustained effort on the physiological level.

Based on the comment of reviewers n ° 1 concerning the too rapid treatment of the cardiovascular part of exercise limitations, these results have been refined compared to the old version. (cf. Reviewer 1, comment 7)

Comment 7:

I found the section on incremental protocol badly written and quite abrupt. The description of the incremental protocol was vague, and it was not clear why there were 3 separate protocols and what determined which protocol was used. Who decides which protocol to use and why?

Reply 7:

We took note of this remark and tried to explain the different protocols in a more pleasant way for the reader.

The decision to choose the incremental protocol is based on the characteristics of the patient as well as the clinical examination preceding the CPET. The doctor evaluates the theoretical maximum power that should be reached by the patient and then can apply an adaptation according to the feelings of the clinical state and the examination of the patient as described by Wasserman et al.

For example, the protocols used are in this case 5, 10 or 15 watts per minute. Which is in line with the 2018 recommendations of the Perioperative Exercise Testing and Training Society (British Journal of Anaesthesia, 120 (3): 484e500)

The literature reports different evaluation methods depending on the teams. As an example, here is the list of protocols used in the studies included in the 2007 Benzo Meta-analysis. You will see that the teams use increments ranging from 10 to 20 watts per minute. It seemed to us as far as possible relevant to compare statistically that a more "difficult" increment protocol does not seem to be more able to demonstrate a more reliable VO2peak in our population of patients with NSCLC. waiting for surgery.

Study	Incremental protocol
Bechard (1987)	12.5 W/min
Bolliger (1995)	warm-up 20 W, ramp 20 W/min

Boysen (1990)	Treadmill speed began at 2 mph with zero grade. At l-minute intervals, the speed was increased by 1 mph.
Brunelli (2002)	No CPET => estimate of VO2 from work (in watts), itself already estimated on the basis of an equation allowing to estimate work in watts according to the number of steps climbed
Brutsche	warm-up 20 W, ramp 20-Wmin
Epstein (1993)	"work rate determined using the method of Wasserman et al."
Larsen (1997)	10 or 15 W/min
Markos (1989)	100 kilopondmeters/min (***around 16,3 watts/min***)
Morice (1992)	10 W/min
Smith (1984)	10 W/min
Torchio	Only abstract no information
Villani	Only abstract no information
Wang 1999	a standardized work rate (5, 10, 15, or 20 W/min) based on the patient's mass, height, age, and FEV1, as described by Wasserman et al.
Wang 2000	warm-up 15 W, ramp 15/Wmin

Changes in the text:

- **Line 139:** "The selected increment was individualized (between 5 to 15 watts/min according to the predicted maximal capacity and estimated physical level during the consultation preceding the CPET"
- **Lines 237 to 249:** "The incremental protocols were set at 5, 10 or 15 watts per minute, according to the predicted maximal capacity and estimated physical level to theoretically last 8 to 12 minutes. The median load during the three minutes warm-up was 20 (10 - 30) watts, and the median load increment was 10 (10 - 15) watts per minute. The mean duration of the incremental phase (after warm-up) was 5.1 ± 2.0 minutes. A significant, positive correlation was found between the duration of the incremental phase of the CPET and the number of maximality markers achieved: $r = 0.41$, $p<0.0001$. We compared the results obtained by patients who benefited from the three different load increments proposed. The mean duration of the incremental phase was significantly different between the three incremental load protocols (4.3 ± 1.7 min, 5.7 ± 2.0 min, and 5.1 ± 2.6 min for 15, 10 and 5 W/min respectively, overall $p < 0.0001$), and was mainly driven by a significant difference between 15 and 10 W/min increments ($p < 0.001$) (Table 4)."

Comment 8:

The discussion section was well-written and structured. There was good effort made at discussing possible explanations for the findings of this study, with suggestions for potential strategies to reduce the incorrect estimation of surgical risk.

Reply 8: We thank you for this comment.

Comment 9:

In summary, this study showed an important and significant finding in that majority of patients who had undergone CPET in the authors' institution did not achieve their real VO2 max. Unfortunately, it is greatly disadvantaged by the lack of clinical correlation of this high-risk group. In my opinion, it is equally important not to underestimate the reasons why these patients did not achieve maximal effort during CPET. In our department, sometimes we use CPET as a simulator of surgical stress. It is important for patients to engage and be motivated. Unmotivated patients, even at CPET assessment, may have an impact on their post-operative recovery and outcomes. They may not be compliant with physiotherapy and rehabilitation programmes on the ward, therefore increasing the risks of post-operative complications.

Reply 9: We totally agree with this comment. It is possible that VO2peak, whether it is associated with maximal or submaximal effort, remains truly predictive of the risk of complications because it precisely integrates the patient's engagement in their therapeutic plan. Based on this remark we have added a few sentences following the limitations.

Changes in the text:

- **Lines 377 to 383:** "Finally, as suggested by some authors "clinical V̇O2peak is part of the deal" (27). It is possible that although the raw value of V̇O2 was underestimated, this peak value includes the painful or motivational component, and remains predictive of perioperative complications. Unmotivated or painful patients, even at CPET assessment, may have an impact on their post-operative recovery and outcomes. To date, this question remains unanswered and future works are needed to further refine our assessment methods in an increasingly large and severe population of candidates for pulmonary surgery."