Dear Editor and Reviewers

Please receive our grateful thanks for the valuable comments from the Editors and the reviewers regarding our manuscript (No. 40339).

Practically all changes suggested by the editors are accepted and implemented in the manuscript, and can be seen in the "track changes" version of the paper as well as some other minor changes from the authors themselves.

The review-reports are also carefully read and we have changed the article accordingly. A point-by-point summary of all the changes made are described below. We have copy-pasted the reviewers comments and our answers are highlighted in yellow.

We are looking forward to your final decision and hope you will deem the revised manuscript acceptable for publication in World Journal of Transplantation.

To the Editor: The names and the order of the authors are ; Katrine Rolid, Arne K. Andreassen, Marianne Yardley, Elisabeth Bjørkelund, Kristjan Karason, Julia P. Wigh, Christian H. Dall, Finn Gustafsson, Lars Gullestad, Kari Nytrøen. All the authors affiliations are rearranged and all postal codes have been added. The ORCID numbers for the second author Arne K. Andreassen and the sixth author Julia P. Wigh are now filled in in the revised manuscript file, unfortunately we did not have the ORCID numbers of these two authors by the time of the initial submission, so they were not electronically registrated on the submission website. The names of these two co-authors are *not* included in the new copyright license agreement from the editorial office following the revision. We have therefore uploaded the correct copyright license agreement signed by all the *ten* authors from the first submission. If we receive a new copyright license agreement with the names of *all* the ten authors of the manuscript, we will of course sign this agreement and forward it as soon as possible.

Information regarding grant numbers, Consort 2010 statement, Informed consent statement, Biostatistics and Conflict-of-interest statement are filled out and uploaded as pdf files. The main aim is specified (page 4) and a file with an audio core tip is added on page 5. Article highlights are written on page 16-18.

In correspondence with the biostatistician there are made some changes to the statistical section at page 10 and in the tables (1-3). In the univariate regression analyses we decided to use the standardized beta value instead of the unstandardized B, and we added the unadjusted R^2 for all the univariate analyses. In the acknowledgements two more names (Andreas Lundberg Zachrisson and Stefan Grau) are added. According to the manuscript guidelines for revising manuscripts, the tables are moved to the end of the manuscript following the figures. For your information, one abstract with data from this study was presented on the International Society for Heart and Lung Transplantation (ISHLT) 37th Annual Meeting and Scientific Sessions in San Diego 2017 (published in the abstract supplement of *The Journal of Heart and Lung Transplantation* 2017; 36: S176) and two abstracts were presented at the ISHLT 38th Annual Meeting and Scientific Sessions in Nice 2018 (published in the abstract supplement of *The Journal of Heart and Lung Transplantation* 2018; 37: S296 and S196).

One of our co-authors (Yardley, M) is the daughter of a native English speaking father and is fluent in English, we have therefore not achieved a non-native English certificate. If further proofreading of our manuscript is preferred we will be happy to do so.

Reviewer`s code 02726701 comments:
Comments on Clinical features and determinants of VO2peak in de novo heart transplant recipients Introduction Nice introduction. It is concise and clear, but, dismisses that some interested readers could not understand what are the central and peripheral factors determining O2 consumption.

This is now clarified (page 6).

At first glance, it is not easy to imagine why is not the transplanted heart itself the most important factor, I suggest to add a brief explanation that, at the same time, could be useful to understand the manuscript aim that follows: "Determine clinical, hemodynamic end peripheral factors that contributes to explain the reduced exercise capacity". Are the clinical factors not related to hemodynamics?

This is also clarified (page 6).

Material and Methods It is confusing that 72 from 155 patients were excluded because they did not meet the I/E criteria considering that they are very broad: "Clinically stable HTx recipients approximately 8-12 weeks after HTx; Age > 18 years; Both sexes; Receiving immunosuppressive therapy according to local protocols; Patient willing and able to give written informed consent for study participation, and motivated to participate in the study for nine months". I presume that all were adult HTx patients receiving immunosuppression. As the authors explain, measuring oxygen consumption (VO2 peak) is part of the standard postoperative care of HTx patients in the participating hospitals, so it is improable that the patients did not consent to undergo the cardiopulmonary exercise test (CPET). Could it be that some patients did not participate because they were too ill? If this explanation is true, maybe the real VO2 peak measurements could be much worse than 19.4 ml/Kg/min.

This is elaborated and discussed in the limitations section (page 15). All patients included in the study gave their written informed consent. We have now explained in more detail that this study is based on the baseline data of an ongoing randomized controlled study (page 7).

Which were the differences between recruited and not recruited patients?

We have now included reasons for exclusion (page 7), but we do not have sufficient data on the excluded patients to perform between-group statistics on baseline characteristics between excluded and included patients.

The tables state that "63-81" patients (first columns) conform the two study groups. Those number differ from the 72 patients recruited. In the Abstract section it appears 81 as the studied sample.

This seems as a misunderstanding. 72 of the 155 patients assessed for eligibility were excluded. 83 were included, and two of these were excluded before baseline-testing (n=81) For some variables there are some missing data, so the number 63-81 illustrates the amount of missing data and a varying «n». This data is now corrected to 55-81 in table 1. One of the variables had n = 55. The other "Ns" vary from 63-81.

Measurements: This section begins with: "The primary endpoint, VO2peak". Is VO2 peak and outcome or a measurement? Consider that this variable comes for a measurement and, for this same reason, it is an independent variable. The "secondary endpoints" are several

other independent variables that could influence or modulate the "primary endpoint" or main independent variable. All these are confusing, especially considering that the redaction of this subsection seem to describe a next multivariate analysis. Were all measurement performed in the same opportunity with VO2 peak? QoL questionnaires too?

As this study is part of a larger RCT, which now is described in more detail, the primary and secondary outcomes should now be clearer to the readers. All baseline tests were done at the same time (over a 2-day period for each patient). Mean (SD) time post HTx for the baseline-tests was 11 (2) weeks.

Statistical analysis: The authors used several statistical tools after dividing their samples in two halves (bellow or above the mean). Most of the time this kind of analysis compares first and last tertiles or quartiles. Are Cohen´s kappa statistic as high to, in fact, differentiate exactly those patients having VO2 peaks 19.3 vs 19.5 ml/Kg/min if they perform a CPET twice?

Cohen´s kappa statistics have not been used because several exercise tests on each patient was not performed, as it would put too much strain on the patients. Patients with values right around the median value could of course belong to either group - wrong or right. Only one patient had a median value between 19.3 mL/kg/min and 19.5 mL/kg/min and five patients had a median value between 19.1 mL/kg/min and 19.9 mL/kg/min. As only a few patients had VO2 values close to the median value, it is not likely that they have appreciable affected the mean VO2peak value of each group.

From this subsection, it is clear that VO2 peak could be considered a binary primary endpoint after concluding a proper logistic analysis. Authors state: "To identify the degree of association with VO2peak, all relevant variables with $P < 0.05$", It seems that they test everything!! If this is true, p-values must be accordingly adjusted for multiple comparisons to be considered significant.

Thank you for underscoring this. We have of course not tested everything and this needed clarification (page 10). Table 1 shows which variables we have performed univariate regression on. Well-known, previously documented predictors (such as age, ejection fraction, BMI and muscle strength) were of course also evaluated for inclusion in the multiple regression analysis, with or without a significant p-value from univariate regression. As our hypothesis was that both peripheral and central factors have an impact on the variance in VO2peak, we selected the given set of variables to be analyzed in the multiple regression analyses, which is the final model presented here. We chose to present the adjusted R square value in order to adjust for the number of variables included in the model.

Results The description is clear and concise. Low VO2 peak patients are sicker than the high VO2 peak ones. The most interesting finding in this section is those poor performance patients seem to have lower left ventricular ejection fraction, higher NT-pro-BNP and to use lower doses of immunosuppressive agents: Were differences in heart rejection rates between both groups?

No, as shown in table 1 there were no differences between the groups regarding rejections.

As pre transplantation demographic and clinical characteristics are comparable in both groups, most findings could be explained by post-operative events and immunological injuries after HTx are examples of those events. From Table 1 exploration: Do authors think that the two study groups are as comparable as their p-values suggest or could it be that the small study sample preclude to find real existing differences (type 2 error)?

We can of course not omit that any type 2 errors exist, but in most HTx literature, samples are small, and thus, a group of 81 in this setting is not so small after all. Furthermore, for most of the key variables, the data was normally distributed and the performed statistics was carefully checked for underlying assumptions.

From Table 1: Is this figure correct? LVEDD (cm) 4.9 ± 0.5 4.9 ± 0.5 4.9 ± 0.4 0.996 0.39 [0.19, 0.59], <0.001

Yes, we have now checked the data/numbers again, and they are correct.

I did not find the description of Table 3.

A figure legend has now been added to Table 3.

Nevertheless, the multivariate analysis resulted that the predictors of above or below the mean VO2peak are two variables resulting from the same CPET (O2 pulse and HR reserve) and from muscular exercise capacity. Do authors think that these findings could disincentive clinicians and physiotherapist to perform CPETs, even this test been safe?

Results from multiple regression analyses are never «black or white» and should always be interpreted with caution. Our results here suggest merely that early $VO_{2\ peak}$ values are determined by both central and peripheral factors which was our initial hypothesis. Furthermore, O2 pulse is a controversial surrogate for stroke volume, so a better measure of this (echocardiography measured during peak exercise) would have been preferable, but this is difficult to measure and may not give reliable measurements either (We tried this in a previous study, but could only measure during sub-maximal exercise: https://www.ncbi.nlm.nih.gov/pubmed/23185084)
We are a bit unsure what is meant by «…findings could disincentive performance of CPET», but we do not think that these findings in any way will frighten health personnel from performing CPET test on their HTx patients. After decades of exercise testing research at our center, our key message is that this test still is under-utilized and we try to motivate and commend to increased used of CPET because it gives so much useful information: both regarding future prognosis and survival and for individually tailoring of exercise programs and rehabilitation. The safety aspect is extremely important, and should always be given much attention. We have added a paragraph about this in the manuscript (page 15). In short, in our many studies involving CPET in HTx recipients, we have never had any serious adverse events, neither in maintenance nor in de novo recipients. As new knowledge now continuously adds to the existing database, the conclusion inclines towards CPET being safe, when closely supervised and performed by competent and experienced health personnel.

Graphs are nice and clear. Discussion This section is round! Well and clearly written. I suggest to include my above commentaries to this section and to think if some of the findings could be explained because most patients were heart transplanted too late. Abstract and Core

tip sections are both OK. In summary. The manuscript is not yet fitted to be published, but, if some modifications are wisely added, it will.

Regarding that some patients may be transplanted too late, we could not see any relationship between years with heart failure prior to HTx and VO_{2peak} as shown in table 1. Thank you very much for all your valuable suggestions. Based on your comments we have added paragraphs and sentences both to the limitation (page 15) and discussion sections (page 15).

Reviewer`s code 03291363 comments:

The major concern I have is the study design. There is no comparator or control group to compare the findings in de novo heart transplants. These controls could be the general population and/ or maintenance heart transplants.

Thank you for your important comments.
It certainly would have been valuable with a control group, but this is baseline data from a RCT study, which is now clarified on page 7. We did not have resources to include a third arm with a healthy population as a control group. We do compare the de novo HTx population with an earlier study from our center (page 13-14) and with other studies (page 12-14), as well as with reference material from a general population (page 10 and 12).

Second there is no longitudinal data to establish the validity of the time selected in the post op course of the recipients.

As stated earlier, this is baseline data from an ongoing RCT, so no longitudinal data is presented in this paper.

Finally I think that the authors should state the hypothesis clearly at the outset in the paper.

We have now clarified our hypothesis in the outset of our paper (page 6).

Hence given the above it is difficult to draw the conclusions of the study namely that there are three determinants of the cardiorespiratory exercise tolerance in de novo heart transplant recipients.

We agree, results from a multiple regression analysis on data from one time-point only (baseline) conclusions must of course be interpreted with caution. However, we believe these data on 81 de novo HTx recipients add valuable knowledge about factors predicting early $VO_{2\ peak}$.

But the paper is well written, the graphs and tables clearly presented and the discussion relevant although too long. A lot of work has gone into this study.The topic itself is important in the care of these heart recipients post op.

Thank you very much!

Reviewer`s code 04382473 comments:
A well done study, but with limited scope. The concept of frailty is increasingly being recognized as a factor influencing long term renal transplant outcomes. while true that Vo2 measurements helps frame expectations of functional capacity post heart transplants, the much more interesting question in my mind is -outcomes. Does a reduced VO2 max predict increasing risk of heart failure, death, rejection, sepsis and/or infection? these remain to hopefully be seen in future publications.

Thank you so much for your comments.
Our research group has previously published a paper demonstrating the relationship between VO_{2peak} and survival: https://www.ncbi.nlm.nih.gov/pubmed/26589579

The publication is of professional quality-and I will recommend publication. Though many interesting variables remain to be explored from the population you are following linking functional exercise testing with outcomes would be a high quality publication needed in heart transplant literature.

We agree, long term follow up of this population is very important. Both a 1- and 3-yr follow-up of this current cohort will be published later.

Reviewer`s code 03742333 comments:
I have read with great interest the study entitled "Clinical features and determinants of VO2peak in de novo heart transplant recipients". Initially, I would like to congratulate the authors for the efforts to develop this multicentre study. The paper is well written and add relevant new data in the field. The authors examine a cohort of 81 patients after de novo heart transplant and following a thoughtful analysis conclude that cardiac (O2 pulse and HR reserve) and peripheral factors (muscular exercise capacity) predict the peak of oxygen uptake in an early stage. The study has merit however I truly believe that some important issues should be addressed. General comments: The number for a clinical trial is provided. According with this number the study is described as an interventional randomized clinical trial and has as interventions moderate and high intensity interval training. However, the abstract starts already saying that this is "cross-sectional analysis". If I am not wrong the data from a clinical trial was used in this observational study? This is an important point that the authors need to clarify as there are huge differences between randomized clinical trials and observational studies. A cross-sectional study, as the study is classified in the abstract, applies one test at a set point in time and assess the prevalence of one factor. However, as it does not have an evolution in time, any causal relationship should be carefully assessed. If this is really the case, this is a limitation that should be acknowledged in the discussion.

Thank you very much for your comments. Everything you question has now been addressed as a response to the other reviewers´ comments above. It should now be much clearer that this paper reports baseline characteristics and findings from an ongoing RCT.

Accordingly, the CONSORT 2010 statement is said to be reported. However there is not a single answer in the form uploaded.

This was a mistake and the CONSORT statement is re-uploaded with answers.

Therefore, if the authors intend to report the results of a clinical trial many elements are missing, such as clear definition of intervention groups, clear end-points of the study and inclusion/exclusion criteria. Was the study randomized? Which technique of randomization was used?

A design paper of the RCT is previously published and we now refer to this study (ref. 22). This current study´s main aim was to split the total baseline cohort in two and compare those with low capacity with those with high capacity. Thus, the focus here is not the two intervention-arms, and we believe it would only be confusing to point this out in this publication.

Title: This is a journal of transplantation in general, I would consider replace the abbreviation VO2 peak by peak oxygen consumption as it is easily understandable for non-cardiac transplant professionals.

We agree, but as we are only allowed to have 12 words in the title, that is why we have used the abbreviation. Instead, we now use peak oxygen consumption in the introduction (page 6).

Abstracts: According with the former comment, the aim could contain some background to contextualise readers about the importance of the subject and VO2 peak and HTx.

We would like to add some more information, but we are only allowed to have 20 words in the aim.

Method section has a wrong typo "perfomed mean 11" and also should explain how experimental groups were defined.

The typo is corrected. We have also defined the experimental groups in the method section of the abstract.

Results mention a general population that was not defined as included in the study. I would wonder if this is a third group?

The general population mentioned is not a third group, but is based on data from reference material from the literature. We have more clearly defined the reference population on page 12 .

Moreover, how psychosocial function was assessed? Is this relevant to be included in the abstract?

Psychosocial function was assessed by Short form-36. It may be not necessary to include this in the abstract as it is a secondary measure. It is now removed from the abstract.

Conclusion says "central and peripheral factors" however these factors were not described previously.

We have now mentioned these factors under "method" in the abstract.

Introduction: This section would benefit from more clarification of what is the importance of VO2 peak and define which the central and peripheral factors are.

This has now been clarified on page (6).

Methods: All general comments apply in here. Was this an interventional clinical trial? How sample size was determined? What was the end-point used to power the study?

As earlier mentioned, this is a baseline study from an ongoing RCT and this is now clarified on page 7.

Results: Really long table 1 making difficult the interpretation. Is it possible to split the table by different topics?

Yes, the table is quite long. We have split table 1 by different topics.

The n in the table varies from 63 to 81, this means that more than 20% of the population is missing for some analysis. This is a factor that can significantly affect results, maybe one column with the n used in each analysis could make readers aware of limitations of some analysis.

For some variables there are some missing data, but we think if we put the "n" in every analysis, the tables will be quite complex to read. The "Ns" are now reported in the multiple regression, table 3.

Discussion: It starts saying "as compared with a general population". I cannot find a third group compared in the study. Therefore I assume you compared with figure reported by previous studies, this should be better explained to avoid confusion.

You are right. There is no third group and we compare our findings with earlier studies. We have now clarified this on page 12.

Is there any limitation that readers should be aware? Please comment.

There are of course a number of limitations. A "limitations" section is now added on page 15.