Author’s response to reviews

Title: Size-adjusted muscle power and muscle metabolism in patients with cystic fibrosis are equal to healthy controls – a case control study

Authors:

Katharina Ruf (ruf_k@ukw.de)
Meinrad Beer (Meinrad.Beer@uniklinik-ulm.de)
Herbert Köstler (Koestler_H@ukw.de)
Andreas Max Weng (weng_a@ukw.de)
Henning Neubauer (henning.neubauer@srh.de)
Alexander Klein (alexander_klein@outlook.com)
Kathleen Platek (KPlatek@t-online.de)
Kristina Roth (roth_kl@ukw.de)
Ralph Beneke (ralph.beneke@staff.uni-marburg.de)
Helge Hebestreit (hebestreit_h@ukw.de)

Version: 1 Date: 07 Oct 2019

Author’s response to reviews:

Editor Comments:

Thank you for submitting your manuscript to BMC Pulmonary Medicine.

As you will see it has been thoroughly peer-reviewed.

I am pleased to invite you to submit a revised manuscript for consideration - please note that the detailed comments will need to be dealt with satisfactorily in a major revision.
The authors thank the editor for the opportunity to revise this manuscript. This has been done according to the reviewers’ suggestions.

Reviewer reports:

Mathieu Gruet (Reviewer 1): I read with great interest the paper by Ruf et al. dealing with muscle function in people with cystic fibrosis (CF).

There is a current debate among researchers and clinicians in the field of muscle and CF around the question of an intrinsic skeletal muscle dysfunction in CF. This debate takes its origins about twenty years ago but has gained a renewed interest with the discovery few years ago of CFTR expression in the animals and humans skeletal muscles (e.g. Divangahi et al, Plos Genet 2009; Lamhonwah et al. Annals of neurology 2010). Expression of a deficient CFTR in the skeletal muscles of people with CF could lead to several negative scenario leading for instance to ionic homeostasis dysregulation, dysregulation of production/utilization of some metabolites (e.g. PCr), that may finally induce early/exaggerated muscle fatigability and exercise intolerance. In a clinical perspective, if such scenarios are true, it means that it will be harder to fully restore skeletal muscle function of these patients, even if correcting for the usual muscle atrophy observed in these patients.

Regarding the literature, some studies suggest the existence of intrinsic metabolic abnormalities in CF (e.g. Erickson et al. Exp Physiol 2015; de Meer et al. Thorax 1995, Selvadurai AJRCCM 2003), whereas other do not (e.g. Werkman et al. Exp Physiol 2016; Decorte et al. JCF 2017). Such problematic has been discussed elsewhere (J Physiol, crosstalk debate 2017; Gruet et al. JCF 2017).

One important methodological consideration is to control for the potential confounding factors related to muscle deconditioning (lack of physical activity) and the associated atrophy. In other terms, it is important to determine if the skeletal muscle abnormalities in CF are independent of muscle atrophy.
The present study by Ruf et al. brings further elements regarding this topic and suggests the absence of a specific deficit when adjusting between group differences for muscle size, which is smaller in the CF population. More specifically, differences in performance and/or metabolism during a (i) Wingate test, (ii) a quadriceps incremental test with low frequency and long stages and (iii) a high-intensity constant load quadriceps exercise, all disappear when adjusting for muscle size.

These results are in accordance with some previous studies (e.g. Gruet et al. JCF 2016; Decorte et al. JCF 2017) which found no differences in neuromuscular function (e.g. muscle contractility, sarcolemmal excitability) and metabolism (e.g. PCr, ADP, pH, Pi/PCr) between mild-moderate CF patients and healthy controls matched for physical activity levels. However, it is contradictory to other studies (e.g. Selvadurai AJRCCM 2003).

The main novelty of the study by Ruf et al. is to operate a statistical analysis that allows to directly control for the influence of some potential confounders, with an ANCOVA considering quadriceps cross-sectional area and height of the subjects.

Overall, this is a well-designed study, this paper is well written, the analysis is robust, and the message that cross-sectional area is really important to consider when comparing muscle performance is supported by the results of the three different tests.

However, some methodological considerations must be considered and some complementary analyses may be necessary to fully support the present results and the conclusion of this study. Moreover, I also have some suggestions to further improve the introduction and discussion.

We thank the reviewer for this thorough and dedicated review of our manuscript. Please find all the changes marked in the manuscript and the respective answers to your comments below.
* Introduction:

In the general background, the authors mixed findings regarding reduced aerobic capacity and those regarding reduced anaerobic power. When stating for instance that exercise limitations in CF are related to pulmonary disease and other factors such as inflammation, it is unclear if the authors speak in general or rather speak about anaerobic or aerobic capacity. Moreover, the clinical implications of altered skeletal muscle function is not enough developed here and then it is not clear why it is important to focus on skeletal muscle function in this population.

I suggest to slightly reformulate the first paragraph considering the following elements:

1- CF is associated with poor exercise tolerance which has major clinical consequences (e.g. survival, QoL, etc…).

We thank the reviewer for this comment and have elaborated on this in the introduction.

2 - Exercise intolerance concerns both aerobic and anaerobic function.

2A - Several factors can explain reduced aerobic capacity in this population, including ventilatory mechanics and gas exchanges abnormalities. However, it seems that peripheral muscle function is also important to consider (many possibilities here to justify the importance of skeletal muscle function during aerobic exercise in CF: for instance, bronchodilators, despite improving lung function, failed to improve peak aerobic capacity in people with CF, e.g. Dodd et al. JCF 2005, or, symptoms of muscle effort evaluated by RPE scale during incremental cycling exercise are equivalent in intensity to symptoms of breathlessness, and that whatever the intensity of exercise (Gruet et al. EJAP 2018), confirming that despite their important ventilatory limitations, patients with CF are still able to fatigue their locomotor muscles).

2B - Muscle function is also important for anaerobic performance (e.g. all the papers that demonstrated correlations between fat free mass and performance during Wingate test in CF).
We thank the reviewer for these valuable remarks. The mentioned references have been incorporated into the description of muscle function in aerobic exercise. The introduction has been rephrased to emphasize the differentiation of aerobic and anaerobic function and give the reader a better understanding of the complexity of muscle function in CF.

3 - Thus, muscle function is a potential important limiting factor of performance in CF, which justify why we must shed light on its specific role. The natural question that arises is to determine whether it is just a quantitative problem (e.g. reduced muscle mass) or whether it is also a qualitative problem (i.e. directly linked to CFTR expression in skeletal muscles and the resultant metabolic abnormalities).

In this context, it is important to investigate the skeletal muscle function in isolation (reducing the potential confounding role of respiratory factors in alteration of muscle metabolism). It is in my opinion worth mentioning, because it is a limitation of studies which investigate muscle metabolism (for instance using NIRS) during whole body exercise. On the other hands, it should be noted that one limitation of using local contractions is the limited ecological validity of the findings [regarding this issue, did you try to correlate some changes in metabolites during the quadriceps exercise, or simply the performance during the quadriceps test (e.g. maximal workload), with the performance during the Wingate test? One strength of your study is that you have, in addition to local exercise tests, a whole body test that is one of the gold standard to evaluate muscle power. Try to establish a link between alterations in local exercise metabolism with a more functional muscle testing (Wingate) is interesting and, to my knowledge, has not been previously reported. It would permit to reinforce the transfer of your mechanistic results to real-word settings.]

We thank the reviewer for this important comment and rephrased the introduction to further clarify our approach.

4 - In this context, local exercise using MRI is interesting for several reasons. Previous studies found conflicting resulted, as illustrated by the recent crosstalk in J Physiol 2017 "Skeletal muscle oxidative capacity is/is not altered in cystic fibrosis patients". One way to clarify this important question is to adjust for variations in body size and muscle mass, which has not been adequately performed in previous studies.
We thank the reviewer for this comment and agree that existing research shows conflicting results. The respective citations were included in this rephrased paragraph where the role of adjustment has been further explained.

I thus suggest to slightly reformulate the introduction considering those elements.

We thank the reviewer for this thorough revision of the introduction, which has been completely revised according to the suggestions.

Then, page 5 L56: the authors stated that "only very few studies have adequately adjusted for differences in body size when muscle function was evaluated. In most studies, either no adjustment was employed at all or a ratio to body weight was calculated. This latter approach has been proven to be inadequate [19]."

I think that this should be better elaborated, especially since the novelty of the present study mostly relies on this methodological aspect. I understand the idea that the scaling was not optimal in previous studies. However, I am not sure that we can state that no adjustment has been employed in previous studies, or just body weight was considered. For instance, in Decorte et al., JCF 2017 subjects performed calf exercise with MRI during incremental exercise using absolute increments, and found that the differences disappeared when normalizing for differences in calf size. They also used an exercise at 50% maximal workload performed for 12 minutes. In Gruet et al. JCF 2016, they performed a quadriceps incremental exercise based on a % of maximal voluntary contraction (MVC). By using such exercises based on relative force level, you can compare your groups because differences in strength (and thus muscle mass) will not directly influence the performance. While I understand that scaling using ANCOVA is important and probably better that using a ratio or an exercise performed at a relative force level (as it is used in most endurance / fatigue studies), I believe this should be better elaborated and justified. You also state that the approach which consists to use a ratio of body weight is not appropriate and cite Tanner JAP 1949. In support from this old reference, I think that more detailed explanations are necessary here, especially because the ratio of body weight is something very popular among clinical exercise physiologists. For instance, VO2peak is always expressed in absolute value (e.g mL/min) but also relative to weight (i.e. mL/min/kg). Thus, further explanations here will be very useful for the readers interested by controlling their outcome for body weight / muscle mass, etc…
We thank the reviewer for this comment and have included a small discussion on scaling in the introduction, including the mentioned references. We think this further clarifies our approach of using an ANCOVA.

* Methods:

- I am not sure to understand how and why you used increments ranging for 0.5 to 1 kg during the first quadriceps test. The increment may vary up to a factor of two, which is not frequent when considering usual ways to increment exercise (it would be similar for instance to have increments during cycling CPET that vary from, 30 to 60 watts). You stated that this is based on patient's anticipated maximal workload. How do you quantify / make this prediction? This is an important issue as it may influence the achievement of the last workload, especially because you used very long stages.

We thank the reviewer for this comment and try to explain our rationale here: To our knowledge, this is the first incremental exercise test of this kind in an MRI in CF. The maximum workload was estimated as follows: male: 0.12 kg load per kg body weight; female: 0.09 kg load per kg body weight. The start load was 50-60% of the estimated maximal workload whichever was rounded to the next 0.5 kg. The increments of 0.5 of 1 kg were chosen in a way that participants had an overall exercise time of about 45 minutes while the duration of stages was 5 minutes (this duration is explained further down as response to your next comment). This worked fairly well for our group of participants as can be seen in the exercise time and the achieved maximal workload. The mean exercise time has been added to the patients’ characteristics table for clarification.
From an exercise-physiology point of view, it can also be explained the following way:

Maximal muscle power of a static muscle contraction is somewhere between 40 and 100 N, depending on training status of the individual (since we did not have trained athletes, we chose to calculate with 40 N). In a dynamic contraction, maximal muscle power equals about 70-80% of the latter. Therefore, 40 N x 0.7 (for the dynamic contraction) x 0.33 (since this is an incremental test and not a single bout of exercise) equals 10N/cm2 muscle cross sectional area. When we assume a quadriceps cross sectional area of 63m2 (which is the mean of the control group of this study) maximal power equals 630 N. Since we need to consider the distance of knee and force transducer (about 30 cm) as well as knee joint and patella (about 4 cm), the ratio of lever arm compared to work arm is 1/7.5. Therefore, the resulting power at the force transducer is about 8.4 kg (630N/7.5 = 84 N~ 8.4 kg). The calculated maximal power was 8.2 kg, therefore the methodological and exercise-physiological considerations both worked well in our cohort to determine maximal power and to determine the increments to achieve a mean exercise time of about 45 minutes in all participants. Further, no significant differences between the groups were found when comparing the load of increments in a t-Test, which has been added to table 2.

- I am very interested by your protocol and the use of (very) long stages. This is a different approach from what it is usually performed in the literature (most of the time it is 1-3min stages). I think it may have an interest to obtain a steady state for some metabolic parameters? This may be further justified here and maybe discussed as a potential novelty as compared to all previous protocols that used shorter stages. One negative aspect of such very long protocol with absolute increments which are not directly based on maximal force or power, is the likely high variability in time to exhaustion. Whereas such exercise may mainly rely on aerobic metabolism in some subjects, it may be different in other subjects for which the first increments are already very hard to perform. Could you please add some informations regarding individual data for time to exhaustion (in both CF and controls) ?

We agree with the reviewer that the long stages are somewhat unusual and explain this rationale here further:

1) MRI technique requires a certain time to generate measurements, in our case about 60 seconds to gather enough data for averaging to get good quality data

2) As you already assumed, we wanted to obtain a steady state in all subjects. To prevent a high variability in time to exhaustion, we used different increments in relation to the estimated maximal load as explained above. We assumed that healthy participants needed about 2-3 minutes to achieve a steady state. Due to pulmonary constraints in patients with CF we added extra time to enable a steady state.
We understand that this is a novel approach. In order to explain this approach to the reader, this has been added to the methods section. As mentioned above, we integrated the time of exhaustion as well as the mean increment in the manuscript for clarification. We did not find any differences between CF and control with regard to time to exhaustion nor weight of increments which seems to justify this approach.

- You state that your design is interesting because you can compare your CF and controls for exercise performed at a similar level. However, despite mentioning that you recorded metabolites at the end of each 5-min exercise bouts, you only report results at (i) pre exercise, at (ii) maximal load and at (iii) recovery. Why not reporting kinetics of changes at isotime at different time points of your protocol? It ensures optimal comparison for instance, after the completion of a given number of stages, or by normalizing your data in percentage of total time duration (e.g. 25%, 50%, etc…). Without such analyses, you only have a comparison with your last stage at exhaustion, which does not occur at the same time point for all subjects. Moreover, especially in fragile patients and by using very long stages, we can expect an influence of motivational factors that may impact the measurements at exhaustion; That is why it is always interesting to consider other time points early in exercise, to be free of this problem of differences in time to exhaustion and the confounding factor of motivation.

We thank the reviewer for this valuable comment which was taken into account.

Due to the study design, where we tried to have a similar exercise time for all, we hope to have achieved a maximal exhaustion during the last stage in all participants. Of course, as you mention, motivational factors will impact these measurements. Therefore, as suggested, we analysed the results of submaximal data and added this analysis to the manuscript. We averaged the submaximal data in percentage of maximal load (40-49%, 50-59%, 60-69%, 70-79%, 80-89% and 90-99%). Neither for pH nor for Pi/PCr did we find any significant differences. This data is displayed in figure 3.
Regarding the subgroup of patients that perform a second quadriceps test at a higher intensity. It is a very good idea to perform such additional test and I agree with the rationale of the authors for doing this. I also perfectly understand why, from a logistical/feasibility point of view, it was complicated to perform it at the whole group level. However, while you control for potential gender differences for your whole group with an approximate comparable number of female in both groups, you have a strong difference with your subgroup (20% of female in the CF vs 60% in the control group). Previous studies suggest differences in endurance / fatigability etc… between men and females and it is always useful to control for this parameter. In CF, it is also possible that females, for differences reasons, may exhibit greater physiological alterations during exercise than male even after considering lower muscle mass (although it is just a hypothesis, but see for instance the particular results of Selvadurai et al. AJRCCM 2003 which found significant metabolic alterations in CF females compared to CF controls, despite similar muscle mass and greater level of habitual physical activity in CF females). Maybe this issue should be underlined as a potential limitation.

We thank the reviewer for this valuable comment and agree that the number of females is low in the subgroup. We are aware of the existing data on different exercise capacity in men and women. Since the number of participants in the subgroup is low, it was statistically not possible nor reasonable to do a gender-separate analysis or consider gender as confounder. We added a statement on this issue to the limitations section.

- Why using only an 8-min recovery period between the two quadriceps tests? It is possible that all subjects did not start the second quadriceps test with a full muscle recovery and inter-individuals differences in fitness and ability to recover from an exercise conducted until exhaustion may have slightly influenced the results of the second test.

When the study was designed, we expected to see larger changes in muscle metabolism already at relatively low exercise intensities than ultimately observed (Figure 3) and considered to look also at recovery kinetics. However, the moderate changes observed precluded such an approach. Since no significant difference was observed between groups at the beginning of the high-intensity exercise, although some participants may not have fully recovered, we believe that recovery was sufficient enough not to have altered results at peak exercise.
Furthermore, technical reasons of MRI-imaging did not allow much longer breaks between exercise tests.

Regarding all these comments, the discussion should be edited to better reflect these methodological considerations and further underline the novelty of the study compared to previous studies.

We thank the reviewer for all these valuable comments. We have elaborated on the methodological considerations in the introduction as well as in the discussion to further emphasize these aspects.

* Tables: why using the word "sprint" in Table 4

The word “sprint” has been deleted.

* Figures: maybe it is just on my computer but the resolution of the figure 1 showing the ergometer is bad. Is it possible to reupload with an improved quality?

We apologize for the bad resolution of the images and reuploaded these with improved quality.
Jana DeBrandt (Reviewer 2): Manuscript - General comment: please be consistent in wording: sometimes you use incremental cardiopulmonary cycling test, sometimes incremental cycling test. Please always use the same wording for the same thing. Also you use the wording 'muscle function', 'muscle power', 'muscle performance', 'muscular exercise capacity', 'muscle capacity'. Please be consistent and accurate to the definitions of each parameter. I agree that with a wingate test you measure power output. But with a knee-extension incremental protocol, I don't think this is power. If you need to maintain a specific load for 5 min (contract - release), I believe this is muscle endurance. Please adapt the whole manuscript based on what you feel are the most appropriate terms but be consistent.

We thank the reviewer for the valuable comment on wording and changed the manuscript according to the suggestions. We decided to use the term “muscle function”, since we think this best reflects what this paper is about, well knowing that sometimes we measure power, sometimes rather endurance. Overall, though, we think muscle function is the best umbrella term to summarize the different exercise tests in this manuscript.

Abstract - Methods - line 16-20: now you state twice that the incremental CPET is performed. I would limit this to one time. Also I don't know if I completely agree with the fact that the incremental CPET is a muscle function test. In part yes, but it is a general exercise test. Also in table 1 you don't put the CPET data under muscle function. You describe them in table 1 under exercise capacity.

We thank the reviewer for this comment which has been considered. The abstract has been reformulated according to your suggestion. We agree that CPET is not necessarily a muscle function test, still, muscle function influences exercise capacity to some extent. In addition to the Wingate and the muscle tests in the MRI we think this may further help explain muscle function but also help describe our population in terms of exercise impairment especially in the group of patients with CF.

Abstract - Methods: you don't mention that qCSA is measured. As this is the outcome you correct your data for, it is very important to incorporate this in your abstract and how you measured it.
We thank the reviewer for this remark and added qCSA to the methods section of the abstract.

Abstract - Results: you don't mention the results on the incremental CPET on the bicycle and muscle metabolism after adjustment.

We thank the reviewer for this comment and added the CPET data to the results section of the abstract. Further, we rephrased this section with emphasis on the linear model used to make it more understandable.

Abstract - Conclusion: you use the term power for all outcomes (see comment above about using the correct wording for you muscle outcomes).

We agree with the reviewer that the wording may be confusion and rephrased this paragraph.

Background - line 10 page 5: change 'poor muscle mass' into 'decreased muscle mass' if that is what you mean.

The respective wording has been changed to your suggestion.

Steroid use and inflammation they play a role in limiting exercise or in initiating muscle dysfunction?

We thank the reviewer for this comment and apologize that this has not further been explained in the manuscript. Steroid use has been shown to decrease muscle power (see e.g. Barry, SC and Gallagher, CG. https://doi.org/10.1152/japplphysiol.00506.2002. Similarly, inflammation in patients with CF is associated with reduced muscle strength. Further, inflammation is associated with lower body weight, also reflecting lower muscle mass (see e.g. Wood, LG et al. J Am Coll Nutr. 2001 Apr;20(2 Suppl):157-65 and Gruet M, et al. J Cyst Fibros 2017, 16(5):538-552). The respective passage has been rephrased to clarify this aspect.
Background - line 18 - page 5: what does the CFTR do in the muscle (also in the discussion you only cover this shortly)?

We thank the reviewer for this important comment and try to explain here: As far as we know, the role of CFTR for muscle function is not yet clear. There has been speculation that a defect in CFTR affects muscle function, however, no causal relationship has been established yet (see e.g. Lamhonwah AM et al. Ann Neurol. 2010 Jun;67(6):802-8. doi: 10.1002/ana.21982. or Divangahi M., PLoS Genet. 2009 Jul;5(7):e1000586. doi: 10.1371/journal.pgen.1000586. Epub 2009 Jul 31). In our research, we try to focus on the functional assessment of the muscle. Certainly, further research, i.e. in CFTR knock-out mouse models might help clarify the role of CFTR for muscle function. We assumed that since we did not find any difference between patients with CF and controls that could not be attributed to scaling, apparently CFTR in the muscle does not relevantly influence muscle. We added a sentence to the results section that hopefully adds to clarify that all differences observed between the groups were of quantitative muscle function (i.e. muscle size) and not qualitative muscle function (i.e. muscle metabolism).

I actually do struggle a bit with the first 2 paragraphs. You say already on line 8-10 that the disease plays a role. Why do you start line 16 with 'However.... the muscle is primarily affected by the disease as...... '. So disease has a major impact apparently. I just don't understand the use of the word 'However'.

We thank the reviewer for this comment and rephrased this paragraph.


We thank the reviewer for this remark. The European guideline for the diagnosis of CF, on which diagnosis of our patients relies, has been added as reference to this section.
Groups were not matched on gender. Do you think this can bias your results?

We agree with the reviewer that this is a drawback of our cohort. We added a remark on the gender aspect to the limitations section. However, for the CPET, predicted values are gender-related. In the linear model, qCSA was entered, which indirectly takes the smaller muscle size of females into account. pH and Pi/PCr are not influenced by gender as far as we know.

Methods - Population: how and from where were participants recruited?

We thank the reviewer for this remark and apologize for the missing information.

Patients were recruited from the local CF clinic, controls were friends of CF patients, hospital staff or friends of the hospital staff. The recruitment details have been specified in the methods section.

Methods - Lung function: FEV1 always needs to have the 1 in subscript.

This has been changed according to the reviewer’s suggestion.

How was diffusion capacity measured. Which technique? Add reference.

We thank the reviewer for this remark: Diffusion capacity was assessed measuring the diffusion capacity for carbon monoxide in the single breath technique (10 s breathhold). A respective reference has been added (Stanojevic S. et al. Eur Respir J 2017, 50(3)).
Methods - exercise testing: how did the familiarization go? Please provide details.

Most patients with CF were familiar with CPET and Wingate, since these are commonly done in our unit. The cycling tasks and the pulmonary function testing were explained (e.g. the duration of stages, the increments, the monitoring) and all the equipment was demonstrated.

After taking place on the bike, patients pedalled for about 20 seconds without load to see if adjustment of crank arms and saddle were comfortable. During the establishment of monitoring (ECG cables, oxygen saturations) patients again received explanations on the upcoming tasks. Upcoming questions were answered at that time.

Before starting the test in the MRI, equipment was demonstrated, the task explained. Participants lay in prone position and the leg was positioned on the coil and fastened with the help of Velcro straps. Then, participants performed 5 repetitions without load before the test was started to get to know the exact task. This also served to make sure that the leg was securely positioned to prevent displacement from the coil during exercise. If necessary, adjustment was done after this try out.

We added this information as requested to the methods section.

How much time was there between Wingate and incremental CPET?

We thank the reviewer for this remark. Patients had a break of at least 30 minutes between these tests to reach recovery. This approached has previously been published and can be found here, for example: Kriemler S, et al.. J Cyst Fibros. 2013 Dec;12(6):714-20. This information has been added to the text.
I think reference 22 is not adequate enough to make sure that the incremental CPET that you are describing is reproducible. I would like to see information on the following: did you measure ECG, blood pressure, saturation? did you have a resting period, warm up, recovery? If yes, how long and at which workload? At how many RPM needed the patient to cycle?

We thank the reviewer for this remark and added further references to describe a maximal effort during the CPET. ECG and oxygen saturation have been continuously monitored during the test, peak heart rate and oxygen saturation at peak exercise have now been added to table 1 to further characterize the effort. RPM was 60 at all stages, this has been added to the methods section.

As explained as response to your above comment, we had a resting period in the beginning that served for the establishment of monitoring and adjustment of the saddle. Before the beginning of the exercise, resting heart rate and SpO2 were monitored. After reaching peak exercise, participants were asked to keep pedalling without load for 2 minutes, however, recovery data, especially gas exchange data was not monitored since this was not the goal of this study.

How was gas exchange measured? Was this Breath by Breath? Was there standardized encouragement during the exercise tests?

We thank the reviewer for this comment and included that gas exchange was measured breath by breath in the methods section. We did not have strictly standardized encouragement, but all tests were performed by the same investigator giving verbal encouragement throughout the test.

Participants between the age of 12 to 42 were included. The protocol of Godfrey is based on childrens height. I guess you did not use this protocol for adolescents and adults? Which protocol did you use for them?

We thank the reviewer for this remark. All participants underwent the Godfrey protocol, regardless of their age. This protocol is widely used to assess patients with Cystic fibrosis and has repeatedly been used for testing adults (see for example Nixon et al. N Engl J Med 1992; 327:1785-1788 or Hebestreit et al. European Respiratory Journal 2006 28: 734-739) and is recommended by the ECFS Exercise Working Group also for adults (Hebestreit et al. 2015).
Methods - MRI spectroscopy:

A) why do you need the steady state exercise over 5 min with a constant load? You don't report results about that part. I don't see the added value. Or do you measure something during those 5 min steady state that is needed for the incremental protocol?

B) How long did the participants needed to let hold the contraction? 1 sec hold, 2 seconds relax? What was the frequency of extensions per minute?

A) Steady state exercise: We thank the reviewer for this comment and understand that this is somewhat misunderstable. When the study was designed, we expected to see larger changes in muscle metabolism already at relatively low exercise intensities than ultimately observed (see the newly incorporated Figure 3 on submaximal data for further illustration) and considered to look also at recovery kinetics. However, the moderate changes observed precluded such an approach. Therefore, this data does not add any further information with regard to our hypothesis.

B) We thank the reviewer for this comment. Participants did not hold the contractions but repetitively moved the leg in a frequency of 2 seconds (i.e. 30/minute). This is elaborated in the methods section. The steady state exercise was done for, adjustment of loads for subsequent exercise and collect recovery data.

Methods: MRI spectroscopy: how did you anticipate the patients maximal load?

We thank the reviewer for this comment and elaborate on this aspect: The maximum workload was estimated as follows: male: 0.12 kg load per kg body weight; female: 0.09 kg load per kg body weight. The start load was 50-60% of the estimated maximal workload, which was rounded to the next 0.5 kg. The increments of 0.5 or 1 kg were chosen to result in an exercise time of about 45 minutes in all participants.
When looking at the increments and the exercise time between the two groups, no significant differences were found.

Methods: MRI spectroscopy: how did you choose the subgroup for the high frequency steady state protocol?

The subgroup was again recruited from the local CF clinic. Actually, the high-frequency protocol was added during the process of the study when we realized that there were no differences between groups with the long stages protocol. We hypothesized that this might be due to the exercise protocol and therefore added the high intensity protocol to reach greater muscle exertion. Patients who were not initially included in the study were asked to participate. Again, friends of patients, hospital staff and their friends were recruited as control group.

Methods: statistical analysis: did you perform a power calculation for the trial? And also for the subgroup trial? Definitely for the subgroup trial 5 vs 10 might be underpowered. Was all data normally distributed? Why didn't you take 'weight' as a control variable as well? This parameter is also different between the 2 groups. Wouldn't it also be interesting to take gender as a control variable as percentages of women in each group are different. Did you test this significantly if this percentage is different?

We thank the reviewer for this comment and try to explain why analyses were done the way we did them. No power analysis has been performed in advance. We are aware that the number of patients included is rather small which is also due to the fact that time in the MRI is scarce and we were happy to include that many participants. Although this is a small number of subjects, we think that we were well able to elaborate on muscle function and muscle metabolism by using a) an MRI-based muscle metabolism measurement, b) a localized muscle test of the quadriceps during the MRI-based metabolism measurements, c) an incremental CPET, d) a Wingate anaerobic test. All these tests were statistically controlled for by measured quadriceps size.

All data included in t-testing or linear models is normally distributed. A statement on this has been included in the methods section.
Since we wanted to control for muscle size, we did not control for weight but rather for the muscle cross-sectional area of the M. quadriceps (qCSA) and height. We believe that qCSA plus height (combined an approximate measure of muscle volume) better represents muscle mass than weight. (An ANCOVA with weight instead of qCSA revealed similar results, still we think that qCSA better reflects muscle mass, therefore this data is not shown). Further, as you mentioned, the number of subjects is rather small and might even be underpowered.

Controlling for a further confounder such as weight, which is, to some extent, included in qCSA, was statistically not feasible. This is already stated in the statistics section of the manuscript. Due to the small sample size, a gender-separate analysis or controlling for gender as separate confounder was neither possible. Still, we agree with the reviewer that gender differences may inflict our results. This issue has now been acknowledged in the limitations section.

Results: do you have data on the length of the testing protocol under MRI per participant? And how much was the increase of maximal load during the incremental knee extension protocol (maximal load minus start load)?

We thank the reviewer for this important remark and apologize for not reporting the test length. This data has now been included in table 2. We did not find any difference in the length of the testing protocol for CF or control.

The increments of 0.5 or 1 kg, respectively, were chosen to enable participants an overall exercise time of about 45 minutes. We started with 50-60% of maximal load, therefore, the increase was about 50% of maximal load in each participant (CF 3.6±1.4 kg, CON 4.3±1.3 kg, p=0.86). If the reviewer feels strongly about it, we can add this data to the manuscript, since it is about half of the maximal load as initially planned, we refrained from doing so at the moment. We did not find any difference in the mean of increments nor in the mean of maximal load between the two groups.

Results: Wouldn't it be interesting to compare the subgroups with the whole cohort also statistically? To see if they differ or not in characteristics? I would add table 4 to table 1 so that you can easily compare the subgroups with the 2 cohorts.
We thank the reviewer for this remark. All patients included in the subgroup are also part of the total cohort in all tests but the high-intensity MRI task. When comparing the subgroup to the rest of the group significant differences were evident in Vo2peak%predicted and in TLCO but not in any other variable. Especially with regard to height and qCSA which were potential confounders and the outcome data of Wingate test and maximal load of the incremental MRI task as outcome parameters for muscle function no differences between the groups were found. Therefore, we do not have evidence of potential differences between the groups that could bias our results.

In our opinion, adding the characteristics of the subgroup to table 1 may lead to a confusing amount of data. Therefore, we added a separate table (table 5) where this data is displayed.

Results - line 34 page 9: please use VO2 peak with '2' always in subscript instead of VO2 max

This has been changed according to your suggestion.

Results: please add peak workload data if available to incremental CPET data. Did you also control the incremental CPET data for qCSA and height. I would be interested to see what comes out of that analysis. In the discussion you state that the difference remains after controlling. If you state this, you need to write this in your results as well.

We have now added maximal oxygen uptake/qCSA to table 1 and added a sentence on this result to the results section.

Results: I miss data on the steady state knee extension test. Why is this data not reported?
We thank the reviewer for this remark. As explained above in the methods section and response to your comment, the initial idea was to analyse recovery kinetics after the first steady state. We expected to see larger changes in muscle metabolism already at relatively low exercise intensities than ultimately observed (see Figure 3 on submaximal data for further illustration). However, the only moderate changes observed precluded such an approach. Therefore, this data does not add any further information with regard to our hypothesis.

In addition to the metabolic data during maximal effort and recovery, we now included submaximal data in the results section to further clarify this aspect. We did not find any significant differences of pH or Pi/PCr ratio between the groups at 40-49%, 50-59%, 60-69%, 70-79%, 80-89% and 90-99% of maximal work load.

If the reviewer feels strongly about it, we can delete the steady state part from the methods section.

Results: According to the correlations. Now you perform the correlation for the whole group (CON and CF together). Is it not more appropriate to perform the correlations separately per group? Now you mix your population, which makes it difficult to interpret the data correctly.

We thank the reviewer for this question and try to shed some light on this aspect. Although separate analyses for CON and CF are reasonable in most aspects, we do not think that a separate analysis will help to further understand the relations between qCSA and anaerobic capacity. Since the Wingate test focuses on anaerobic capacity, decrements in lung function will merely influence the outcome of this test. When calculating correlations separate for the groups, results are as follows: quadriceps-peak power CF r=0.638 (p=0.01), CON r=0.826 (p=0.00), quadriceps-mean power CF r=0.826 (p=0.00), CON r=0.824 (p=0.00), quadriceps-power drop CF r=0.474 (p=0.022), CON r=0.560 (p=0.010). For further clarification of this aspect, we included next to the correlation of the total group the separate correlations of CF and CON into Figure 2.
Results: high intensity protocol. I think you need to be careful with this analysis. This is only a 5 vs 10 analysis and I guess the study is not powered for that sample size. If you look at table 4, you see that the MRI spectroscopy data actually is different between the groups. And this was not like this in the incremental knee extension protocol. Don't you think because of the small sample, you might not reach statistical significance while actually there seems to be a different response. Can you test more patients with this protocol to create more power for this analysis? And can you also balance the gender out better, now it is 1 vs 6 women.

We agree with the reviewer that the sample size of this study is small which impairs some conclusions. As elaborated above, MRI spectroscopy is an expensive analysis and time in the MRI is scarce. When looking at the subgroup data, the only - non-significant - difference seen is Pi/PCr at peak exercise, neither in pH nor in the recovery data any difference can be observed. For sure, a larger cohort may lead to further information, still this remains speculation. We will not be able to perform any further testing at the moment. We added a comment on the sample size and gender-separate analysis to the limitations section to take this issue into account.

Results: muscle metabolism: is the paired sample t-test performed in the whole cohort? CON and CF together? If so, please make this clear. Would it be interesting to perform a paired sample t-test per group?

We thank the reviewer for this question and try to explain the statistical methods further:

The t-tests were performed on the subgroup that performed both the incremental and the high-intensity exercise task to underline that the high-intensity protocol actually leads to a greater exertion. For the question, whether a change of protocol leads to greater exertion the differentiation of CF and CON is not relevant, especially since we were able to show that neither in the incremental protocol nor in the high intensity protocol significant differences between CON and CF were evident. Further, the linear model showed that disease status did not influence muscle function. We rephrased this passage of the results section slightly to clarify that the whole subgroup was included in this analysis.

Results: are the MRI spectroscopy data also controlled for qCSA and height?
As explained in the methods section, spectroscopy data was gathered by using a localized voxel with a defined size which was the same for all patients. From MRI spectroscopy, we do not report absolute values but ratios (iP/PCr) or shifts of peaks (pH). No size related adjustments are required.

Discussion - line 30 page 11: what do you mean with 'no qualitative functional differences to the leg muscles of healthy controls was observed'. Why qualitative? You have quantitative data.

We thank the reviewer for this comment and apologize for this misleading wording. We have elaborated on the difference of qualitative and quantitative muscle function in the introduction. In this manuscript, we see “qualitative” as a primary involvement of muscle function, whereas a “quantitative” is regarded as a difference of muscle mass.

Discussion - line 60 page 11: i think you need to write 'the aforementioned potential....'

This section has been rewritten to your proposal.

Discussion: I believe that only the left leg was used in the testing protocol. Did you question the participants on sport activities? Maybe it was better to use the dominant or the non-dominant leg because sport activities can influence that.

We thank the reviewer for this interesting remark. The MRI-exercise task was an easy, non-complex movement and for reasons of standardization we decided to use the left leg in all participants. Also, positioning of the coil in the MRI was standardized by using the same leg in each participant. Of course, training my influence exercise performance, even in such an easy motor task. Still, we think that the amount of influence on the test outcome is but marginal.

Discussion: line 33 - 38 page 13: Physical activity is not the same als exercise training. I don't think by controlling for QCSA that you also control for training status. Please remove this from the discussion.
We thank the reviewer for this remark and apologize for this confusion. This has been removed from the discussion.

Tables - general comment: please be consistent in use of points, plus-minus, etc. Now there is sometimes no space, too much spaces, small letters, big letters, subscripts, etc....

Tables have been reviewed according to your suggestions.

Tables: table legend: just write ***p<0.001, ** p<0.01, * p<0.05 instead of the way you have done it now (please adapt for every table).

This has been changed as suggested.

Tables: Add % of women in row 1 of tables.

This has been changed as suggested.