PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Strength training as a supplemental therapy for androgen deficiency of the aging male (ADAM): Study protocol for a three-arm clinical trial.
AUTHORS	Kralik, Michal; Cvecka, Jan; Buzgo, Gabriel; Putala, Matus; Ukropcova, Barbara; Ukropec, Jozef; Killinger, Zdenko; Payer, Juraj; Kollarik, Boris; Bujdak, Peter; Raastad, Truls; Sedliak, Milan

VERSION 1 – REVIEW

REVIEWER	gustavo monnerat federal university of rio de janeiro
REVIEW RETURNED	04-Sep-2018

GENERAL COMMENTS	Kralik et colleagues presented an interesting study proposal named: “Strength training as a supplemental therapy of androgen deficiency of the aging male (ADAM): rationale and design”. This study is very interesting and can open novel data regarding key aspects of male health and response to hormone treatment and exercise, two main strategies for health and well-being. Comments: Key point: the statistics analysis is not described in the manuscript (Statistical methods for analyzing primary and secondary outcomes), neither the method for sample size calculation in order to estimated number of participants and blinding for the data analysis. The percentages of causes of hypogonadism in the patients is not clear. Minor: the author considers the training volume appropriate? Some recent publication demonstrated the need o higher volume. At least some discussion about this would be interesting.

REVIEWER	Julius Fink Juntendo University Graduate School of Medicine Japan
REVIEW RETURNED	04-Nov-2018

GENERAL COMMENTS	Very interesting study design. The outcomes of this study might majorly contribute to the current body of evidence in the field of TRT. However, the manuscript needs several minor corrections listed below. Page 8 Line 39 on "the" role of testosterone...Some "report" not "reports", same correction line 42

	Page 9 Line 30 Replace “substitution” with “replacement” Page 9 Line 35 8 year treatment not 8 year Treatment Page 10 Line 3-5 Please be careful with regard to statements of the effects of T on the prostate, you might wanna read the review from Y Cui et al., 2014 Prostate Cancer and Prostatic Diseases Page 10 Line 7-8 Please rephrase: Bashin’s study did not find “high” incidence in old people but “higher” incidence as compared to young people. In fact, frequency of total and serious adverse events and prostate events by testosterone dose was not statistically different between young and old men! Only 9 individuals out of 61 showed severe side effects. Furthermore, from this data you can not say that side effects occurred in the 300-600mg groups only. Page 10 Line 22 has been conducted on “the” effects of … Page 10 Line 25 60-90 s is short to medium duration Page 11 Line 4-5 The addition of resistance training to TRT showed increased T levels compared to TRT alone?! Resistance training induced T elevations last only for a short time after training and do not chronically induce major serum elevations. Please make sure all the information is accurate. You can refer to J Fink et al., The role of hormones in muscle hypertrophy, Phys Sportsmed, 2018. Page 12 Line 3 all groups engaged in, not “are” engaged. Page 12 Line 6-21. Please check grammar. Page 13 Line 9-10. Mostly unilateral exercises? Are those specified in the methods? Why unilateral? I don’t see the reason why. Table 2 1+3 sets? Please explain. Include rest period between sets. Page 11 Line 47 Please add reasoning for n=12 per group. Page 12 Line 18-20 Why undecanoate? Don’t different esters with more frequent injections lead to more stable T levels? Why 1000mg/12 weeks? Blood samples: T should be measured several times during the 12 week period and at different time points to assess the influence of resistance training-induced acute elevations (at rest, immediately after training etc.) Page 14 first paragraph. The addition of upper and lower body muscle CSA measurements via MRI might improve assessment of muscle mass increases. Muscle fiber size and regulators of muscle fiber size. Please specify how you are going to measure satellite cells Please add a detailed statistical analyses section

REVIEWER	Jacqueline K. Dawson, PhD California State University, Long Beach, USA
REVIEW RETURNED	21-Jan-2019

GENERAL COMMENTS	This is in an interesting protocol that proposes to investigate the effect of testosterone replacement therapy (TRT) in combination with strength training on body composition, physical function and cardiometabolic outcomes in hypogonadal males. As the authors have noted, numerous studies have investigated the effect of TRT alone on the above parameters. Thus, the addition of strength training along with the examination of cellular and molecular parameters appears to be a logical next step to investigate and potentially improve physiological outcomes in androgen deficient aging males. I do have some specific concerns regarding several points in the methodology that may affect the clarity of results, which are detailed below. Most importantly, the methods lack a statistical analysis and sample size calculation. Without these sections, it is difficult to ascertain how the study hypotheses (also missing) will be tested. Also, there are grammatical errors throughout the paper (too many to list in this review). The manuscript may benefit from the review of a native English speaker before another submission. Specific Comments: Abstract Missing the Ethics and Dissemination section. I believe the study was registered as a clinical trial, which should be included as a final section in the abstract. Strength and limitations (first bullet, p. 4 line 45) I question the accuracy of this statement as there have been numerous investigations into the effect of exercise training on side effects due to male hypogonadism, particularly if you consider the prostate cancer/androgen deprivation therapy literature. Perhaps focusing on specific aspects of your methodology that make your study novel compared to these investigations would be a better approach. Strength and limitations (second bullet, p. 4 line 51). I would agree that this is a strength. Strength and limitations (third bullet, p. 5 line 3) What is meant by neuromuscular function? I would be more specific here, as neuromuscular function can have different definitions for different people. Also, how is this bullet different from bullet 2? What are the limitations? I would guess that a sample size of n=36 would be underpowered to detect changes in all outcome variables in a 3-arm design and really only demonstrate preliminary insight for future, larger trials. Introduction Overall, the introduction is quite verbose and should be written more succinctly. While there is ample background information

	provided, the introduction reads like a literature review and goes too far in depth on many topics than is necessary. For instance, the section on hypogonadism could be shortened to one paragraph. Instead, focus on building the rationale for your hypothesis and the novelty of your study. For example, why are you interested in examining muscle cellular outcomes? No rationale was given. Also, how do your methods specifically improve on previous protocols? You have a number of outcomes. Which is your primary outcome (i.e. that you would base your sample size calculation on)? Which are your secondary outcomes? What are your hypotheses for the primary/secondary outcomes? Page 5, lines 37-54. This is an example of a paragraph that can be written far more succinctly. Which guideline for hypogonadism are you applying in this study? In figure 1, you state that 6.9 nmol/l is the criterion value which appears to be the AACE guideline. Page 7, lines 52 -52. Should citation [14] be placed at the end of the sentence, or is there another reference for low testosterone levels increasing osteoclast activity? Page 8, lines 39-47. I would encourage the authors to be more precise with their referencing in this section as well as throughout the manuscript. Several statements are made without appropriate supporting references, i.e. “Only a few report strength gains that can be considered substantial in comparison to the benefits of resistance exercise training”. Page 9, lines 16 – 27. These ideas on the effect of TRT on bone density, cognitive function and quality of life need to be more developed. In their current state, they appear as afterthoughts on the paragraph about the effect of TRT on physical function. Page 9, line 17. Reference for this sentence? Page 9-10. Risks associated with TRT It is unclear how this paragraph relates to your methodology, especially since no section on adverse effects were mentioned in the methods. Page 10, lines 53-57. The experimental design of the Glintborg et al. study is unclear. Is it a 2-group design (TRT vs. strength) or something else? Since this study is directly relevant to yours, please describe more completely so the reader can adequately evaluate the literature. Methods and analysis As mentioned in the general comments, a sample size calculation and statistical analyses are missing from this section. Also missing – dates of the study and status of the study (i.e. are you currently in recruitment/data collection?). Inclusion criteria. One inclusion criterion is that participants are hypogonadal. When is this verified in the consent process? Are testosterone levels

	checked prior to consent? Are they verified again during/at the end of the study to ensure levels haven't changed? Page 12, lines 18-21. What is the rationale for giving patients the 1000 mg dose? This should be referenced and/or connected to rationale presented in the introduction. Page 12, line 31. What is the rationale for following the strength training program by Segal et al.? Page 12, line 39. Can you describe the general warmup in greater detail? Table 2. What are the rest periods between sets? Between exercises? Page 14, Body composition. Please describe or add a reference to the method of patient positioning and scanning on the DXA as this can affect reproducibility of results. What are the % CV and % ICC for test-retest reliability? Page 14, lines 26-30. It's stated that leg press 1RM is estimated from multiple repetition maximum testing. Is this the testing that is performed on the first day of training as mentioned in the training protocol design paragraph? If so, it doesn't appear that participants are familiarized with the leg press prior to testing, which can lead to heterogenous results in individuals of different training statuses. Page 14, Muscle strength testing. What are the % CV and % ICC for the MVC and RFD on the dynamometer? Also, what happens if a participant can't perform a test? What is the protocol that is used (i.e. how many practice trials, how many attempts, etc.) Page 14, line 31. What is the outcome measure and protocol for the sit to stand test? Is a stopwatch used? If so, what are the %CV and %ICC? Page 14, lines 31-32. What protocol is followed for the handgrip strength test? Is only the dominant hand tested, or both hands? Is the best out of several trials recorded or just a single trial? Please be more specific. Page 14, lines 35-40. Is the 10 m walk speed measured by a timing gate? What is the protocol that is used? What is the outcome measure? Are multiple trials performed? Page 14, lines 42-47 and Figure 2. Can you indicate on Figure 2 when the familiarization occurs in relation to the outcome assessments? Page 14, Psycho-social functioning. Missing a reference for the Aging Males' Symptom scale. Page 15, Serological outcomes. Please describe where the venous blood sample is obtained (i.e. antecubital fossa) and if it is processed immediately or

	stored/batch tested. If stored, describe processing, storage temperature, and storage duration. How are the samples analyzed? Page 17, Muscle cellular outcomes. What is the rationale for obtaining biopsy samples from only 80% of participants? Also, how is the tissue stored and what is the duration of storage prior to analysis? Finally, it would be quite important to know how long after the last exercise session the second muscle biopsy obtained, as this can affect cellular outcomes. What is the justification for this length of time? General questions or comments on the Methods section. What safety measures are in place for each of the exercise tests and sessions to ensure patients are not injured? What will happen if a participant cannot perform an exercise in the program at the prescribed volume or intensity? How will this be tracked and handled? Do you have alternative exercises? This may relate to program compliance, which is also mentioned below. Compliance to the exercise program can clearly affect study outcomes. How is compliance to the exercise program tracked? Is dietary intake monitored? What instructions are given to participants regarding diet during the intervention? How will exercise/activity performed outside the study be tracked/controlled? Clearly, this can influence study outcomes. Discussion While a discussion section is not required, I wonder if you might use this opportunity to reiterate your study strengths in comparison to the previous literature, as well as discuss study limitations. For example, a major limitation of your study is that it is underpowered to detect difference in the many outcomes you have. Therefore, it should be regarded as a pilot study to provide preliminary insight into a larger, more definitive trial. Authors' contributions I could not find this section in your manuscript. Figure 2. Can you indicate the number of days elapsed on the timeline? This would allow the reader to infer how many days/weeks elapse before the start of the training session and how many days the second biopsy is performed after the post-training assessment.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: gustavo monnerat
Institution and Country: federal university of rio de janeiro
Please state any competing interests or state 'None declared': none

Please leave your comments for the authors below
Kralik et colleagues presented an interesting study proposal named: “Strength training as a supplemental therapy of androgen deficiency of the aging male (ADAM): rationale and design”. This study is very interesting and can open novel data regarding key aspects of male health and response to hormone treatment and exercise, two main strategies for health and well-being.

Comments:
Key point: the statistics analysis is not described in the manuscript (Statistical methods for analysing primary and secondary outcomes), neither the method for sample size calculation in order to estimated number of participants and blinding for the data analysis.
The percentages of causes of hypogonadism in the patients is not clear.

RESPONSE: Firstly, authors would like to thank to the editor and all reviewers for their comments and suggestions, we appreciate the effort and we feel that adjustments made based on your comment improved the manuscript substantially!

Regarding the statistics, the following text has been added:”
Normality of the data distribution will be assessed by comparing histogram of the sample data to a normal probability curve and outliers will be identified as values distant for more than 3σ from the average. Normality will be further tested with Kolmogorov-Smirnov test if needed.
Differences between normally distributed variables will be evaluated by the Analysis of variance with repeated measures and Bonferroni post-hoc test, differences between pre- and post-training values of the specific subpopulation will be evaluated with a paired Student’s t-test. Non-normally distributed variables will be log transformed. Variables that could not be log transformed to normal distribution will be tested with non-parametric tests (Mann-Whitney test and Wilcoxon rank test).
Cohen’s d will be used to calculate effect size (ES), represented by ‘d’ and interpreted as < 0.2 is a small, 0.2-0.8 is a moderate, and > 0.8 is a large effect size.
For studying the relationships between the various outcomes, the Pearson or Spearman correlation tests will be used.
All statistics will be performed using a statistical software and P values < 0.05 will be considered significant. Data will be presented as means and standard deviations.

Regarding the sample size calculation: the pre-existing data from our previous 12 week exercise intervention study related to fat (5.6% decrease p=0.002) & lean body mass (1.8% increase, p=0.047, DEXA) and that of maximal voluntary contraction force measured on linear leg-press (31% increase,p<0.0001, 1RM) - were used to determine sample size for the population of the designed intervention study. The Type I error probability was set at 0.05 and the power to 0.90 and 0.95. Original number of subjects (n=12) was achieved when the power was set to 0.80. Results in table below clearly indicate that 20 patients per group will be sufficient to detect exercise intervention related changes in fat mass, and muscle strength, and that at least 22 patients would be required to detect the difference in lean body mass at the power of 0.90
Therefore, we increased the number of subject per group to 22.

	Actual Power	Type I error	Effect size	Total sample size
Fat mass (units)	0.951	0.05	0.7665	20
Fat mass (units)	0.916	0.05	0.7665	17
Lean body mass (units)	0.956	0.05	0.7326	27
Lean body mass (units)	0.906	0.05	0.7326	22
MVC leg press (1RM)	0.956	0.05	1.3245	8
MVC leg press (1RM)	0.924	0.05	1.3245	7

Regarding the percentages of causes of hypogonadism, we are not sure if we understand it correctly: Do you mean the percentage of primary vs. secondary hypogonadism cases in our recruited subjects. At the moment, 99,6 % of the recruited subjects had primary hypogonadism, 0,4 % of the cases are unclear.

Minor: the author considers the training volume appropriate? Some recent publication demonstrated the need of higher volume. At least some discussion about this would be interesting.

RESPONSE: Thank you for the comment, we made internal decision that, based on our previous experiences with similar groups of patients with low testosterone training for changes in body composition, which are the primary outcomes, training twice a week should be an adequate frequency given the total load in each session. Most studies in this type of patients under strength training intervention used shorter session of the strength training part (20 – 40 minutes), lower number of set 2-3 or lower volume of the repetitions 6 – 10. Since the participants are almost novices in a mean of strength training, with no or very little history of strength training in the past, we decided to use lower number of training sessions, partly also for a safety reasons and to avoid overreaching of the participants.

We also added additional information to the text as follows: “The strength training protocol will be a modified strength exercise program from Segal et al. [55] which was used in similar patients as ours. The participants will perform 24 training sessions of strength training protocol with the frequency of two training sessions per week for 12 weeks. There will be at least 48 hours rest period between two subsequent training sessions (Monday and Thursday). The intervention will take place at the Faculty of Physical Education and Sport, Comenius University in Bratislava, Slovakia. All training sessions will be supervised and guided by professionals with university degree in sports training to ensure safety, correct technique and progression in training load, with a maximum of three participants per one trainer.”

Reviewer: 2
Reviewer Name: Julius Fink
Institution and Country: Juntendo University, Graduate School of Medicine, Japan
Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below
Very interesting study design. The outcomes of this study might majorly contribute to the current body of evidence in the field of TRT.
However, the manuscript needs several minor corrections listed below.

RESPONSE: Firstly, authors would like to thank to the editor and all reviewers for their comments and suggestions, we appreciate the effort and we feel that adjustments made based on your comment improved the manuscript substantially.

Page 8 Line 39 on "the" role of testosterone...Some "report" not "reports", same correction line 42

Page 9 Line 30 Replace “substitution” with “replacement”

Page 9 Line 35 8 year treatment not 8 year Treatment

RESPONSE: Thank you for these language corrections, we changed them in the manuscript accordingly.

Page 10 Line 3-5 Please be careful with regard to statements of the effects of T on the prostate, you might wanna read the review from Y Cui et al., 2014 Prostate Cancer and Prostatic Diseases

RESPONSE: Thank you for the comment, we are aware that the research of the prostate cancer and TRT provide contradictory results at the moment. We decided to erase the prostate cancer from the statement. We changed the paragraph to:

Testosterone treatment is contraindicated in subjects with breast cancer or benign prostate hyperplasia, lower urinary tract symptoms, and if risks of treatment is perceived to be high by many physicians [3]. The risk of prostate cancer is still unclear. Only intra-muscular treatment found slight increase in PSA levels [41]. Loeb et al. [42] found that TRT remained significantly associated with more favourable-risk prostate cancer and lower risk of aggressive prostate cancer. But other studies and meta-analysis found TRT as a safe urological approach to treat hypogonadism [43-44]. Other risks of TRT in men include fluid retention, mood fluctuations, gynecomastia, worsening of sleep apnea, polycythaemia, elevation of PSA [3,5, 45].

Page 10 Line 7-8 Please rephrase: Bashin's study did not find "high" incidence in old people but "higher" incidence as compared to young people. In fact, frequency of total and serious adverse events and prostate events by testosterone dose was not statistically different between young and old men! Only 9 individuals out of 61 showed severe side effects. Furthermore, from this data you can not say that side effects occurred in the 300-600mg groups only.

RESPONSE: Thank you for the comment, we changed the lines of the Bhasin study to: Bhasin et al. [46] found higher incidence of adverse effects (included haematocrit greater than 54%, leg oedema with shortness of breath, urinary retention and prostate cancer) in treating older men with the very high doses of T compared to young males. Rhoden and Morgentaler [47] have reviewed the adverse effects and recommend the long-term monitoring of the above-mentioned parameters. Potential adverse events not related to hormones include pain at injection site and local skin irritation.

Page 10 Line 22 has been conducted on "the" effects of …

Page 10 Line 25 60-90 s is short to medium duration

RESPONSE: Thank you, we include these changes into the manuscript.

Page 11 Line 4-5 The addition of resistance training to TRT showed increased T levels compared to TRT alone?! Resistance training induced T elevations last only for a short time after training and do not chronically induce major serum elevations. Please make sure all the information is accurate. You can refer to J Fink et al., The role of hormones in muscle hypertrophy, Phys Sportsmed, 2018.

RESPONSE: Thank you for your comment, we described the results of the studies more precisely as follows:
"Glintborg et al. [53] studied effects of TRT and/or strength training (ST) on cardiovascular risk in hypogonadal males for 6 months. This double-blinded, placebo-controlled study found that only ST + placebo significantly decreased sCD36 levels. Only placebo group did not decrease fat mass during

this period. Compared to TRT, six months of strength training reduced sCD36 levels suggesting decreased cardiovascular risk, possibly due to a reduction in central fat mass.
In a pilot randomized controlled trial by Cho and colleagues [54] showed, that when hypogonadal males were treated with combination of exercise and TRT, they showed significantly better results in serum testosterone levels and symptoms of hypogonadism compared to TRT alone after 12 weeks of intervention. The levels of testosterone were significantly higher in the combination group ($p = 0.01$) In addition, these improvements were well-maintained in the combination group with continuous exercise even after cessation of TRT. After 20 weeks of intervention the group which used TRT and strength training kept the testosterone levels significantly higher ($p = 0.01$) compared to the group with TRT only. Consequently, it seems that exercise can augment the durability of response to TRT and it may be the solution to shorten the treatment duration with a lower risk from testosterone therapy [54].
There are some very promising results showing a great potential of exercise in hypogonadal patients. However, the above-mentioned studies did not focus on possible physiological and metabolic mechanisms responsible for the positive effects of resistance training at circulating, cellular and molecular level. Up to date, there are no studies investigating the effects of strength training on the regulation of muscle mass and neuromuscular function on a cellular level in hypogonadal male patients."

Page 12 Line 3 all groups engaged in, not "are" engaged.

Page 12 Line 6-21. Please check grammar.

RESPONSE: Thank you, we changed the wording and corrected the grammar issues.

Page 13 Line 9-10. Mostly unilateral exercises? Are those specified in the methods? Why unilateral? I don't see the reason why.

RESPONSE: The design of the intervention was written not clearly, the only unilateral exercise used is "split squat". The reasoning was added to the manuscript as follows:
"Since unilateral exercises (e.g. squats) develop similar magnitude of muscle activity with producing less load on the spine, thus they are safer [56], the split squats are chosen instead of regular squats."

Table 2 1+3 sets? Please explain. Include rest period between sets.

RESPONSE: 1+3 sets was meant to be a warm-up set, but we deleted the 1 from the manuscript, so it would not be confusing for the readers.
Also the rest period were added to table 2 and described in the manuscript as followes:
"The strength protocol exercises will be performed with free weights and on machines. The training program consist of six exercises for upper and lower body at an intensity of 60-80% (8 – 12RM: the load that induces technique failure in eight or twelve repetitions) of one-repetition maximum and takes approximately 60 minutes. The inability to perform full repetition will be assessed by supervisor or rejected by participants. The participants will be instructed to perform a concentric action in 2 s and immediately after an eccentric action in also 2 s. There will be 90 seconds' rest period after each set. The same duration rest period will be between all the exercises. The rest periods will be controlled by timer (The miniMAX, Gymboss, USA). The load will be added, if participant can complete prescribed number of repetition in each set of the exercise. More detailed strength training protocol can be seen in Table 3. During the first three weeks of the intervention, there will be one set in the beginning with light weight to focus on safety and technique. After that three more sets with weight close to 60 – 80 % of 1RM will follow. After first three weeks, there will be increase in number of sets to four."

Page 12 Line 18-20 Why undecanoate? Don't different esters with more frequent injections lead to more stable T levels? Why 1000mg/12 weeks?
RESPONSE: Testosterone undecanoate (Nebido) is the only injectable form of testosterone used at the institutes of collaboration physicians of the study. According our knowledge, this form of T at dose of 1000 mg is the most stable from all available preparations for 3 months period, which is considered a standard treatment in Slovakia. Shorter-acting forms may cause more pronounced fluctuations in 24-hour circulating levels of testosterone.

Blood samples: T should be measured several times during the 12 week period and at different time points to assess the influence of resistance training-induced acute elevations (at rest, immediately after training etc.)

RESPONSE: Thank you for your recommendation, but we decided not to track acute responses, since the actual physiological meanings of acute hormonal responses for muscle mass accretion etc. has been questioned in the recent years by some studies. As to the more frequent testing of resting T values, it could be interesting to follow. However, we do not expect significant changes in resting T due to training.

Page 14 first paragraph. The addition of upper and lower body muscle CSA measurements via MRI might improve assessment of muscle mass increases.

RESPONSE: Thank you for the recommendation; the MRI would certainly add a valuable information. However, we used DXA data only due to financial limitation of the study and also an increased testing burden to the subjects.

Muscle fiber size and regulators of muscle fiber size. Please specify how you are going to measure satellite cells

RESPONSE: Thank you, The number of satellite cells will be quantified on frozen muscle cross sections with a immunohistochemical protocol as described in Bjornsen et al. 2019 (Pax7 + Laminin + DAPI)

Please add a detailed statistical analyses section

RESPONSE: Regarding the statistics, the following text has been added:
Normality of the data distribution will be assessed by comparing histogram of the sample data to a normal probability curve and outliers will be identified as values distant for more than 3σ from the average. Normality will be further tested with Kolmogorov-Smirnov test if needed.
Differences between normally distributed variables will be evaluated by the Analysis of variance with repeated measures and Bonferroni post-hoc test, Differences between pre- and post-training values of the specific subpopulation will be evaluated with a paired Student's t-test. Non-normally distributed variables will be log transformed . Variables that could not be log transformed to normal distribution will be tested with non-parametric tests (Mann-Whitney test and Wilcoxon rank test).
Cohen's d will be used to calculate effect size (ES), represented by 'd' and interpreted as < 0.2 is a small, 0.2-0.8 is a moderate, and > 0.8 is a large effect size.
For studying the relationships between the various outcomes, the Pearson or Spearman correlation tests will be used.
All statistics will be performed using a statistical software and P values < 0.05 will be considered significant. Data will be presented as means and standard deviations.

We also re-calucated the required number of subjects per group. The pre-existing data from our previous 12 week exercise intervention study related to fat (5.6% decrease p=0.002) & lean body mass (1.8% increase, p=0.047, DEXA) and that of maximal voluntary contraction force measured on linear leg-press (31% increase,p<0.0001, 1RM) - were used to determine sample size for the population of the designed intervention study. The Type I error probability was set at 0.05 and the power to 0.90 and 0.95. Original number of subjects (n=12) was achieved when the power was set to 0.80. Results in table below clearly indicate that 20 patients per group will be sufficient to detect exercise intervention related changes in fat mass, and muscle strength, and that at least 22 patients would be required to detect the difference in lean body mass at the power of 0.90
Therefore, we increased the number of subject per group to 22.

	Actual Power	Type I error	Effect size	Total sample size
Fat mass (units)	0.951	0.05	0.7665	20
Fat mass (units)	0.916	0.05	0.7665	17
Lean body mass (units)	0.956	0.05	0.7326	27
Lean body mass (units)	0.906	0.05	0.7326	22
MVC leg press (1RM)	0.956	0.05	1.3245	8
MVC leg press (1RM)	0.924	0.05	1.3245	7

Reviewer: 3
Reviewer Name: Jacqueline K. Dawson, PhD
Institution and Country: California State University, Long Beach, USA
Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below
This is in an interesting protocol that proposes to investigate the effect of testosterone replacement therapy (TRT) in combination with strength training on body composition, physical function and cardiometabolic outcomes in hypogonadal males. As the authors have noted, numerous studies have investigated the effect of TRT alone on the above parameters. Thus, the addition of strength training along with the examination of cellular and molecular parameters appears to be a logical next step to investigate and potentially improve physiological outcomes in androgen deficient aging males.

I do have some specific concerns regarding several points in the methodology that may affect the clarity of results, which are detailed below. Most importantly, the methods lack a statistical analysis and sample size calculation. Without these sections, it is difficult to ascertain how the study hypotheses (also missing) will be tested.

RESPONSE: Firstly, authors would like to thank to the editor and all reviewers for their comments and suggestions, we appreciate the effort and we feel that adjustments made based on your comment improved the manuscript substantially!

Also, there are grammatical errors throughout the paper (too many to list in this review). The manuscript may benefit from the review of a native English speaker before another submission.

RESPONSE: Thank you for taking the time and really checking our manuscript in details. We are thankfull and we tried to include all of the asked comment and tried to change the manuscript so I will fit criteria.

Specific Comments:

Abstract
Missing the Ethics and Dissemination section.

RESPONSE: Thank you for the comment, the section was added to the end of the abstract in this form:
“Ethics and Dissemination: This trial is approved by Ethics Committee of the University Hospital in Bratislava, Slovakia (ref. trial number: 127/2017) and all subjects will be fully informed on the rationale, risks and benefits of the study and sign the written informed consent prior entering the study. Results will be published in peer-reviewed journals, presented in scientific conferences and the disseminated to healthcare professionals, and general public in the form of popular journal articles and lectures. Trial registration: ClinicalTrials.gov: NCT03282682.”

I believe the study was registered as a clinical trial, which should be included as a final section in the abstract.

RESPONSE: Thank you, yes, the study is registered and we also include trial registration detail into the end of the manuscript: “Trial registration: ClinicalTrials.gov: NCT03282682.”

Strength and limitations (first bullet, p. 4 line 45)
I question the accuracy of this statement as there have been numerous investigations into the effect of exercise training on side effects due to male hypogonadism, particularly if you consider the prostate cancer/androgen deprivation therapy literature. Perhaps focusing on specific aspects of your methodology that make your study novel compared to these investigations would be a better approach.
Response: Please, see the response bellow.

Strength and limitations (second bullet, p. 4 line 51). I would agree that this is a strength.
Response: Please, see the response bellow
Strength and limitations (third bullet, p. 5 line 3)
What is meant by neuromuscular function? I would be more specific here, as neuromuscular function can have different definitions for different people. Also, how is this bullet different from bullet 2?

Response: Please, see the response bellow

What are the limitations? I would guess that a sample size of n=36 would be underpowered to detect changes in all outcome variables in a 3-arm design and really only demonstrate preliminary insight for future, larger trials.

RESPONSE: Thank you for your comments, we changed this paragraph as follows:

Strengths and limitations of this study
- To the best of our knowledge this trial represents the first study in hypogonadal males focusing on possible physiological and metabolic mechanisms of strength training at circulating, cellular and molecular level.

- Wide spectrum of clinical outcomes with high-standard methods of assessments (DXA, muscle biopsies).

- The major limitation of this trial is small sample size, caused by limited number of detected patients.

Regarding the sample size calculation, the pre-existing data from our previous 12 week exercise intervention study related to fat (5.6% decrease p=0.002) & lean body mass (1.8% increase, p=0.047, DEXA) and that of maximal voluntary contraction force measured on linear leg-press (31% increase,p<0.0001, 1RM) - were used to determine sample size for the population of the designed intervention study. The Type I error probability was set at 0.05 and the power to 0.90 and 0.95. Original number of subjects (n=12) was achieved when the power was set to 0.80. Results in table below clearly indicate that 20 patients per group will be sufficient to detect exercise intervention related changes in fat mass, and muscle strength, and that at least 22 patients would be required to detect the difference in lean body mass at the power of 0.90
Therefore, we increased the number of subject per group to 22.

	Actual Power	Type I error	Effect size	Total sample size
Fat mass (units)	0.951	0.05	0.7665	20
Fat mass (units)	0.916	0.05	0.7665	17
Lean body mass (units)	0.956	0.05	0.7326	27
Lean body mass (units)	0.906	0.05	0.7326	22
MVC leg press (1RM)	0.956	0.05	1.3245	8
MVC leg press (1RM)	0.924	0.05	1.3245	7

Introduction
Overall, the introduction is quite verbose and should be written more succinctly. While there is ample background information provided, the introduction reads like a literature review and goes too far in depth on many topics than is necessary. For instance, the section on hypogonadism could be shortened to one paragraph. Instead, focus on building the rationale for your hypothesis and the novelty of your study.

RESPONSE: Thank you for your comment; we have shortened the unnecessary parts of text so it would be more concise and easier to read.

For example, why are you interested in examining muscle cellular outcomes? No rationale was given. Also, how do your methods specifically improve on previous protocols?

RESPONSE: Thank you, we feel that adding new information about adaptation at cellular and molecular level would be the next logical step to existing current literature in this area of research. Specifically, whether the satellite cell response is blunted in hypogonadal men and thereby limits the muscle hypertrophy known, as well as whether type I fibers are more affected than type II fibers is not known.
(mechanisms + novel use of this type of measurements)

You have a number of outcomes. Which is your primary outcome (i.e. that you would base your sample size calculation on)? Which are your secondary outcomes?

RESPONSE: Thank you for comment. We added specification of the primary and secondary outcomes as follows:
“Primary outcome measure
Lean mass (LM)

The primary outcome of the study will be the change in lean mass (LM) measured by Dual-energy X-ray Absorptiometry (Hologic fan-beam bone densitometer Discovery QDR series). The changes in lower and upper body LM are investigated separately because of differences in androgen sensitivity in leg muscles compared to neck, chest and shoulder muscles [57]. Due to very similar results but greater participant comfort [58] we decided to use The National Health and Nutrition Examination Survey (NHANES) protocol, which required the participant to be positioned in a supine position in the middle of the densitometry table with head straight, space between the arms and torso, palms flat on the table, and feet together secured by a strap. The study of Shiel et al. [58] in the systematic review showed a strong level of agreement as illustrated by high ICC's and CCC's between the Nana and NHANES positioning protocols, however systematic bias within limit of agreement plot and a large difference in 95% confidence limits indicates that the protocols should not be interchanged when assessing an individual.

Secondary outcome measures
Body composition
Other body composition parameters (fat mass, total body mass) will be measured at the same time so also the protocol is the same as with the primary outcome. The height by stadiometer and waist circumference is measured by stretch-resistant tape that provides a constant 100 g tension. The body mass index is afterwards calculated and reported.
…
“

What are your hypotheses for the primary/secondary outcomes?

RESPONSE: Thank you for your comment; the hypothesis for the primary outcome is that NON-TRT group will gain less lean muscle mass compared to TRT or eugonadal men.
The same will apply to fat mass in means on negative relationship between T level and fat mass changes.
Strength and functional parameters will not be significantly affected with T levels.

Page 5, lines 37-54. This is an example of a paragraph that can be written far more succinctly. Which guideline for hypogonadism are you applying in this study? In figure 1, you state that 6.9 nmol/l is the criterion value which appears to be the AACE guideline.

RESPONSE: Thank you for comment. We follow the AACE through the manuscript and the study, so we change the text accordingly:
“Although the low limit of normal total testosterone is not clearly defined, American Association of Clinical Endocrinologists (AACE) suggests 6.9 nmol/l as lower limit of normal testosterone levels, other societies suggest 8 nmol／l and even up to 10 nmol/l as a limit below which patients can be considered as hypogonadal. The Endocrine Society defines male hypogonadism as a clinical syndrome resulting from failure of the testis to produce physiological levels of testosterone (androgen deficiency) and normal number of spermatozoa. Hypogonadism (primary, secondary or mixed) is caused by disruption of one or more levels of the hypothalamic–pituitary–gonadal axis [7]. All the causes of male hypogonadism can be found in Table 1. Due to complexity of the diagnosis of hypogonadism, there are several alternative names for male hypogonadism, but for a purpose of this trial, the term ADAM was chosen.”

Page 7, lines 52 -52. Should citation [14] be placed at the end of the sentence, or is there another reference for low testosterone levels increasing osteoclast activity?

RESPONSE: Thank you, the sentence was changed so it more obvious that is citation from the same study “Androgens also have a direct impact on bone mineral density since testosterone and oestrogens both play a vital role in bone health and low testosterone levels can cause an increase in osteoclast induced bone resorption [13].”

Page 8, lines 39-47. I would encourage the authors to be more precise with their referencing in this section as well as throughout the manuscript. Several statements are made without appropriate supporting references, i.e. “Only a few report strength gains that can be considered substantial in comparison to the benefits of resistance exercise training”.

RESPONSE: Thank you, the text has been modified so that it includes references to the statements. “Among published trials on the role of testosterone in older men, not all report increased muscle strength with testosterone replacement therapy. The studies reporting significant strength gains were performed in hypogonadal subjects and employed a higher dose of testosterone, for a longer duration [26].”

Page 9, lines 16 – 27. These ideas on the effect of TRT on bone density, cognitive function and quality of life need to be more developed. In their current state, they appear as afterthoughts on the paragraph about the effect of TRT on physical function.

RESPONSE: Thank you for comment. In this trail we are trying to focus mostly on the muscle mass and fat mass. Due to duration of our trial, (12 weeks), we do not expect significant changes in bone density, so we did not try to go more deeply into the issue of bone.
We therefore added more about testosterone and quality of life and cognition.

Page 9, line 17. Reference for this sentence?

RESPONSE: Thank you, the reference was added as follows:
Maintenance of the musculo-skeletal system by increased bone density will contribute to increased physical fitness, reflected by increased strength and endurance [33]. Treatment outcome is strongly influenced by age and training [33].

Page 9-10. Risks associated with TRT
It is unclear how this paragraph relates to your methodology, especially since no section on adverse effects were mentioned in the methods.

RESPONSE: Thank you for your comment. We do not aim to study the adverse effect of TRT, as you pointed. What we tried to emphasizes here is that there are possible problems associated with TRT and that strength training could be used as a part of the possible treatment in situations when TRT is not recommended.

Page 10, lines 53-57.
The experimental design of the Glintborg et al. study is unclear. Is it a 2-group design (TRT vs. strength) or something else? Since this study is directly relevant to yours, please describe more completely so the reader can adequately evaluate the literature.

RESPONSE: Thank you for comment. Two studies were mixed here and therefore we corrected and changed the text and added more details as suggested:
“Glintborg et al. [53] studied effects of TRT and/or strength training (ST) on cardiovascular risk in hypogonadal males for 6 months. This double-blinded, placebo-controlled study found that only ST + placebo significantly decreased sCD36 levels. Only placebo group did not decrease fat mass during this period. Compared to TRT, six months of strength training reduced sCD36 levels suggesting decreased cardiovascular risk, possibly due to a reduction in central fat mass.
In a pilot randomized controlled trial by Cho and colleagues [54] showed, that when hypogonadal males were treated with combination of exercise and TRT, they showed significantly better results in serum testosterone levels and symptoms of hypogonadism compared to TRT alone after 12 weeks of intervention. The levels of testosterone were significantly higher in the combination group ($p = 0.01$) In addition, these improvements were well-maintained in the combination group with continuous exercise even after cessation of TRT. After 20 weeks of intervention the group which used TRT and strength training kept the testosterone levels significantly higher ($p = 0.01$) compared to the group with TRT only. Consequently, it seems that exercise can augment the durability of response to TRT and it may be the solution to shorten the treatment duration with a lower risk from testosterone therapy [54].
There are some very promising results showing a great potential of exercise in hypogonadal patients. However, the above-mentioned studies did not focus on possible physiological and metabolic mechanisms responsible for the positive effects of resistance training at circulating, cellular and molecular level. Up to date, there are no studies investigating the effects of strength training on the regulation of muscle mass and neuromuscular function on a cellular level in hypogonadal male patients.“

Methods and analysis
As mentioned in the general comments, a sample size calculation and statistical analyses are missing from this section.

Also missing – dates of the study and status of the study (i.e. are you currently in recruitment/data collection?).

RESPONSE: Thank you for a comment. We included the trial status in manuscript: “At the time of the first submission of the protocol, the trial was in the phase of participant recruitment. The recruitment began in February 2017 and the last part of data collection is expected to end in August 2019. “

Inclusion criteria.
One inclusion criterion is that participants are hypogonadal. When is this verified in the consent process? Are testosterone levels checked prior to consent? Are they verified again during/at the end of the study to ensure levels haven’t changed?

RESPONSE: Thank you for the comment. The T levels are checked at the meeting with physician a few days before the consent is signed. The T levels are checked also after the training intervention. We included this into the manuscript “The most important inclusion criteria for participation in the study from the patient population are age 40-60 years old, subjects of hypogonadism on testosterone replacement therapy and newly diagnosed patients of hypogonadism. The hypogonadal patients suggested for study participation will be verified for low testosterone, before entering the study. The same verification will take place at the end of the study.”

Page 12, lines 18-21.
What is the rationale for giving patients the 1000 mg dose? This should be referenced and/or connected to rationale presented in the introduction.

RESPONSE: Testosterone undecanoate (Nebido) is the only injectable form of testosterone used at the institutes of collaboration physicians of the study. According our knowledge, this form of T at dose of 1000 mg is the most stable from all available preparations for 3 months period, which is considered a standard treatment in Slovakia. Shorter-acting forms may cause more pronounced fluctuations in 24-hour circulating levels of testosterone.

Page 12, line 31. What is the rationale for following the strength training program by Segal et al.?

RESPONSE: The training program was a modified version of Segal's et al. program originally used on prostate cancer patients on T deprivation therapy and also used in similar group of patients in our labs. Therefore we selected a similar program also for this study, as the current population is also low in T.

Page 12, line 39. Can you describe the general warmup in greater detail?

RESPONSE: Thank you, we added a table describing the warm up protocol:

General dynamic Warmup Exercises	Number of repetitions
Walking low skip	8 times each leg
Walking high knee skip	8 times each leg
Walking knee to chest	8 times each leg
Walking hamstring stretch	6 times each leg
Walking lunge	6 times each leg
Standing lateral lunge	6 times each leg
Egyptian mobility exercise	6 times each arm
External rotation exercise	6 times each arm
Hip hinge exercise	8 times
Air squat	8 times

Table 2. What are the rest periods between sets? Between exercises?

RESPONSE: Thank you for comment. Table 2 was extended with "Rest period" and we also added the following text in the manuscript: "After each set there is a rest period for 90 seconds. The same duration of rest period is applied between the exercises. The rest periods are controlled by The miniMAX timer (Gymboss, USA).".

Page 14, Body composition. Please describe or add a reference to the method of patient positioning and scanning on the DXA as this can affect reproducibility of results.

RESPONSE: Thank you, we added references and described DXA in more details as follow:
The primary outcome of the study will be the change in lean mass (LM) measured by Dual-energy X-ray Absorptiometry (Hologic fan-beam bone densitometer Discovery QDR series). The changes in lower and upper body LM are investigated separately because of differences in androgen sensitivity in leg muscles compared to neck, chest and shoulder muscles [57]. Due to very similar results but greater participant comfort [58] we decided to use The National Health and Nutrition Examination Survey (NHANES) protocol, which required the participant to be positioned in a supine position in the middle of the densitometry table with head straight, space between the arms and torso, palms flat on the table, and feet together secured by a strap. The study of Shiel et al. [58] in the systematic review

showed a strong level of agreement as illustrated by high ICC's and CCC's between the Nana and NHANES positioning protocols, however systematic bias within limit of agreement plot and a large difference in 95% confidence limits indicates that the protocols should not be interchanged when assessing an individual.

Page 14, lines 26-30.
It's stated that leg press 1RM is estimated from multiple repetition maximum testing. Is this the testing that is performed on the first day of training as mentioned in the training protocol design paragraph? If so, it doesn't appear that participants are familiarized with the leg press prior to testing, which can lead to heterogenous results in individuals of different training statuses.

RESPONSE: Thank you for comment. The leg press 1RM is not assessed during the first training session. At the beginning of the intervention protocol, we conduct diagnostics for 10 and 12 RM for all training included exercises: "In the first training session, 10RM and 12RM diagnostic test for all the exercises are conducted."
The testing of the 1RM on leg press is a part of the muscle strength testing prior the start of the intervention, which includes familiarization of the tests one week prior pre-intervention testing.

Page 14, Muscle strength testing.
What are the % CV and % ICC for the MVC and RFD on the dynamometer? Also, what happens if a participant can't perform a test? What is the protocol that is used (i.e. how many practice trials, how many attempts, etc.)

RESPONSE: Thank you for comment, we specified details according your question as follows:
"Muscle strength of lower extremities will be measured by maximal voluntary contraction (MVC) of isometric knee extension and isometric knee flexion by Novel Portable Isometric Knee Dynamometer (ARS dynamometry, S2P ltd., Ljubljana, Slovenia). Each of the test will be performed 6 time with three practise trials and three recorded trials. For the first practise trial, participants will be instructed to achieve approximately 50% of the maximum, with 20 seconds rest period. The second and third practise trial will be performed at 80% of the maximum with 20 second rest periods. The last three trials will be performed with maximal voluntary effort and will be recorded. The best out of three will be taken for further analyses. Rest period during recorded trials will be 60 seconds. During the MVC, the participants will be asked to push/pull as strong as possible and hold for five seconds. Intra-session repeatability for MVC is the 5.7 CV % and 0,98 ICC. Additionally, with awareness of health issues (such as higher blood pressure) and because of a safety reasons, dynamic leg press 1RM (one repetition maximum) will be predicted from multiple repetition maximum testing [59]. "

Page 14, line 31.
What is the outcome measure and protocol for the sit to stand test? Is a stopwatch used? If so, what are the %CV and %ICC?

RESPONSE: Thank you for comment. We decided to erase this test from the study due to low quality of the data and reliability in this subpopulation. This test is more suitable in frail older adults.

Page 14, lines 31-32.
What protocol is followed for the handgrip strength test? Is only the dominant hand tested, or both hands? Is the best out of several trials recorded or just a single trial? Please be more specific.

RESPONSE: Thank you for comment. We added more details for this test as follows: “Muscle strength of the upper extremities, the isometric MVC handgrip strength will be measured by Camry Digital Hand Dynamometer. The participant will stand upright and holds the dynamometer in the hand next to the body, with the minimal or none flexion in the elbow joint. The base of the handle will be on the first metacarpal, while the handle should rest on the middle of the four ringers. None of the body parts will be allowed to move. The test will be performed with three practise trials. First on 50% and the others on 80% of their perceived maximum with 20 seconds’ rest period. After that, three trials with rest period of 60 seconds will be recorded and the best out of three will be marked as the best. The participants will be encouraged to give their maximum effort. The participant will squeeze the dynamometer for 5 seconds. After the test with dominant hand, the test is performed for non-dominant hand. “

Page 14, lines 35-40.
Is the 10 m walk speed measured by a timing gate? What is the protocol that is used? What is the outcome measure? Are multiple trials performed?

RESPONSE: Thank you for comment. We added more details for this test as follows: “10-m preferred walk-speed and 10-m maximum walk-speed will be measured by timing gates WITTY GATE (MicroGate, Italy). Participants will walk 10 meters and the time will be measured for the intermediate 6 meters. This allow acceleration and deceleration. The gates will be placed on 2-meter mark and 8-meter mark. The timing starts when participant cross the first mark and stop when the 8-meter mark is crossed. There will be three trials for preferred and three trials for maximum walk-speed. The outcome measure will be velocity in meters per second calculated as mean of the three trials or the best trial from the preferred and maximum walk-speed test, respectively. Participants will be asked to perform at preferred walking speed first followedand then at the fastest walking speed possible. “

Page 14, lines 42-47 and Figure 2.
Can you indicate on Figure 2 when the familiarization occurs in relation to the outcome assessments?

RESPONSE: Thank you for comment. We modified the figure 2, since the familiarization occurs 7 days before the assessments.

Page 14, Psycho-social functioning.
Missing a reference for the Aging Males’ Symptom scale.

RESPONSE: Thank you for comment, we added the following text:
“The general health status is measured by The Short Form (36) Health Survey patient-reported survey of patient health (SF-36). In addition, clinically investigating the health-related quality of life (HRQoL) symptoms of aging men are measured by Aging Males' Symptom (AMS) Scale. The AMS scale had internal consistency [$\alpha = 0.89$ (95% CI 0.88-0.90)]; the mean alpha estimates across the AMS subscales ranged from 0.79 to 0.82. The AMS scale also had good test-retest reliability [$r = 0.85$ (95% CI 0.82-0.88]; the test-retest reliability coefficients of the AMS subscales ranged from 0.76 to 0.83 [61]. AMS is a standardized scale according to psychometric norms. Most of the currently available language versions were translated following international standards for linguistic and cultural translation of quality of life scales. [62]. “

Page 15, Serological outcomes.

Please describe where the venous blood sample is obtained (i.e. antecubital fossa) and if it is processed immediately or stored/batch tested. If stored, describe processing, storage temperature, and storage duration. How are the samples analyzed?

RESPONSE: Fasting morning venous blood will be taken after overnight (10-hour) fasting and 15 min rest from cubital vein from 8:00 am to 10:00 am [63-64] into closed system collection tubes containing beads coated with a clotting activator and polyacryl ester-gel (Sarstedt AG & Co, Germany). The blood will be centrifuged (3000g, 4°C, 10min) immediately after sampling to obtain EDTA plasma or they will be centrifuged (3000g, 4°C, 20min) after 30min at RT, to obtain serum. The haematological and biochemical parameters analyzed immediately will be haemoglobin, hematocrit, leucocytes, thrombocytes, glucose, urea, sodium, potassium, calcium, ALAT, total cholesterol, LDL cholesterol, HDL cholesterol, triglyceride, testosterone, oestrogen, LH, FSH, SHBG, albumin, bilirubin, total protein, CRP, insulin and PSA. Plasma and serum aliquots (500 ul) will be stored at -20°C (analysis within 6 months) and at -80°C for the long term storage. Bioactive molecules (myokines, exerkines, released from skeletal muscle and/or other tissues) which could be associated with the adaptive response to exercise in all patients will be quantified.

Page 17, Muscle cellular outcomes.
What is the rationale for obtaining biopsy samples from only 80% of participants? Also, how is the tissue stored and what is the duration of storage prior to analysis? Finally, it would be quite important to know how long after the last exercise session the second muscle biopsy obtained, as this can affect cellular outcomes. What is the justification for this length of time?

RESPONSE: Thank you for the comment. According to the Ethical committee requirements, refusing the biopsy sample must not be an exclusion criterion for this study. Although all subjects are asked to provide a biopsy sample, we do not expect 100% compliance. Based on our previous experience, we expect about 80% of the subjects will be willing to undergo this test both before and after the training. The second biopsy is obtained 5 to 7 days after the last training session. This information has been added to the text as follows: “

General questions or comments on the Methods section.

What safety measures are in place for each of the exercise tests and sessions to ensure patients are not injured?

What will happen if a participant cannot perform an exercise in the program at the prescribed volume or intensity? How will this be tracked and handled? Do you have alternative exercises? This may relate to program compliance, which is also mentioned below.

Compliance to the exercise program can clearly affect study outcomes. How is compliance to the exercise program tracked?

RESPONSE: Thank you for these relevant comments. In our study we tried to focus on safety first approach, for that reason, every session will be supervised and guided by professional coaches with university degree in sports training to ensure safety, correct technique and progression in training load, with a maximum of three participants per coach.

We added this information to the text as follows:

After that “The participants will be familiarised with the equipment and exercise technique one week

before the start of the intervention. The technique corrections will be possible during the whole intervention if needed. 10RM and 12RM diagnostic test for all exercises will be conducted during the first week of intervention."

Standard of the full repetition states that repetition is not full when "the load that induces technique failure in eight or twelve repetitions" and "The inability to perform full repetition is judged by supervision or rejected by participants. Each session is managed by at least two professionals, who received strength training programme and record every set and repetition made in each session in training plan", "Prescribed exercises in the strength training protocol for every training sessions can be found in Table 4. Each session will be managed by at least two professionals, who received strength training programme and record every repetition and set made in each session in individual training plan. At the start of each session, the trainers will ask participants if they experienced any adverse events since the last session and record reported events. At the end of the session, the trainers will ask participants if they experienced any adverse events during the session, which will be also recorded. If participant will be unable to perform any of the exercises or sets, this will be recorded into prepared training plan and the situation will be managed during the first week during familiarization with the training protocol and will be solved by the authors. The appropriate alternative exercise will be considered depending on the restriction or participant's limitation.
"

Is dietary intake monitored? What instructions are given to participants regarding diet during the intervention?

RESPONSE: Thank you for your comment. Even though we tried to control as many factors as possible, the idea was not to induce too many changes in volunteers' life during the study. They were therefore instructed not to change their dietary and exercise habits, as it is stated in the manuscript (the statement was recently added) "The participants will be asked to not change their habitual dietary intake and physical activity patterns. "

How will exercise/activity performed outside the study be tracked/controlled? Clearly, this can influence study outcomes.

RESPONSE: Thank you for comment. We will not monitor physical activity levels, which may be a downside of the study bit in the other hand, some studies found monitoring e.g. wrist accelerometers not reliable. The subjects are instructed before and asked about their activities during the intervention. We also added more specific information as follows: "Participants will be asked to continue in physical activities as before, but any kind of regular physical activity, especially strength training or any other kind of weight training during the intervention will be also prohibited. The exclusion criteria reject any participant, who performed any kind of regular supervised strength training one year prior to study.."

Discussion
While a discussion section is not required, I wonder if you might use this opportunity to reiterate your study strengths in comparison to the previous literature, as well as discuss study limitations. For example, a major limitation of your study is that it is underpowered to detect difference in the many outcomes you have. Therefore, it should be regarded as a pilot study to provide preliminary insight into a larger, more definitive trial.

RESPONSE: Thank you for your comment, we agree and changed those sections regarding sample size calculations and strength and limitation as responded above.

Authors' contributions
I could not find this section in your manuscript.

RESPONSE: Thank you for comment, we added this part into the manuscript.
"AUTHORS' CONTRIBUTIONS
MK, JC, TR, MS participated in the study design and drafted the manuscript, GB participated in the development of the intervention protocol, BU and JU designed protocol for biological sample collection and processing, and will participate in biological material sampling & analyses. MP, ZK, JP, BK and PB provide access to patients. MK, JC, JU and MK performed data analysis. All authors contributed to and approved the present manuscript."

Figure 2.
Can you indicate the number of days elapsed on the timeline? This would allow the reader to infer how many days/weeks elapse before the start of the training session and how many days the second biopsy is performed after the post-training assessment.

RESPONSE: Thank you for comment and thoughts. The figure 2 was changed and the weeks are now in each block.

VERSION 2 – REVIEW

REVIEWER	Gustavo Monnerat Federal university of Rio de Janeiro, Brazil
REVIEW RETURNED	05-Apr-2019

GENERAL COMMENTS	no additional comments

REVIEWER	Julius Fink Juntendo University, Graduate School of Medicine, Department of Urology
REVIEW RETURNED	05-Apr-2019

GENERAL COMMENTS	The previous comments have been addressed and the quality of the manuscript has improved. However, several minor corrections should be made: 1) P4 Add references to the 8nmol/l and 10nmol/l suggestions 2) P6 why is it necessary to analyze free testosterone? 3) P6 "If untreated, chronic lower than normal testosterone level dramatically increases risk of many diseases later in life" You could add reference "Mobility and Biomechanical Functions in the Aging Male: Testosterone and the Locomotive Syndrome", Fink et al 2018 4) P7 Do you mean muscle strength or muscle mass, please specify 5) General warm up? What is this based on?

REVIEWER	Jacqueline Dawson, PhD California State University, Long Beach, USA
REVIEW RETURNED	17-Apr-2019

GENERAL COMMENTS	I commend the authors on the revised version of the manuscript. Please see the points below on a few items that are still unclear.

	1) Strength training intervention – p. 14, line 27. “Since unilateral exercise (e.g. squats) …” Do you mean to say split squats? A squat is a bilateral exercise. 2) In the sample size section, it is mentioned that fat, lean mass and MVC on leg press were used to determine the sample size. Please clarify which of these variables were actually used to determine the sample size. Also, it would be helpful to readers to state the value of the change (kg) ± SD for the variable that you are using for the calculation. Finally, how are you accounting for attrition? 3) Please add the expected n to the methods section of the abstract. 4) In the statistical analysis section, a description of how you will manage missing data is absent. Please also list the software that you will use to perform the statistical analysis. 5) How is exercise program and study safety assessed? It is mentioned that familiarization and a qualified instructor will oversee the training, but how are you tracking injuries, adverse events, etc. and how does this affect the calculation of compliance vs. adherence? 6) Is dietary intake monitored? Even though you mention that you are not controlling for diet, tracking dietary intake, whether through a log or other means, can certainly strengthen your protocol. As you’ve already begun data collection, this may not be possible, in which case it should be listed as a limitation of your study design.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:Reviewer: 2Reviewer Name: Julius FinkInstitution and Country: Juntendo University, Graduate School of Medicine, Department of UrologyPlease state any competing interests or state ‘None declared’: NonePlease leave your comments for the authors belowThe previous comments have been addressed and the quality of the manuscript has improved. However, several minor corrections should be made:

1) P4 Add references to the 8nmol/l and 10nmol/l suggestions

RESPONSE: Thank you for the comment, the reference “Wang C et al. Investigation, treatment and monitoring of late-onset hypogonadism in males: ISA, ISSAM, EAU, EAA and ASA recommendations. Eur J Endocrinol. 2008 Nov;159(5):507-14.” was added as follows “Although the lower limit of normal total testosterone is not clearly defined, American Association of Clinical Endocrinologists (AACE) suggests 6.9 nmol/l as lower limit of normal testosterone levels, other societies suggest 8 nmol／l and even up to 10 nmol/l [7] as a limit below which patients can be considered as hypogonadal.”.

2) P6 why is it necessary to analyze free testosterone?

RESPONSE: Thank you for the comment, we added “Most testosterone circulates tightly bound to sex hormone-binding globulin (SHBG) or weakly bound to albumin. A minor amount circulates as free testosterone, and it is believed that this is the metabolically active fraction. Therefore, measurements

of free testosterone is important in the diagnosis of disorders of androgen deficiency in men [16].” with the reference to it.

Shea JL et al. Free testosterone: clinical utility and important analytical aspects of measurement. Adv Clin Chem. 2014;63:59-84.

P6 "If untreated, chronic lower than normal testosterone level dramatically increases risk of many diseases later in life" You could add reference "Mobility and Biomechanical Functions in the Aging Male: Testosterone and the Locomotive Syndrome", Fink et al 2018

1: Fink JE, Hackney AC, Matsumoto M, Maekawa T, Horie S. Mobility and Biomechanical Functions in the Aging Male: Testosterone and the Locomotive Syndrome. Aging Male. 2018 Sep 29:1-8. doi: 10.1080/13685538.2018.1504914. [Epub ahead of print] PubMed PMID: 30269622.

RESPONSE: Thank you, it is nice piece of work, we added it into the main document.

“Testosterone is a hormone regulating several pathways affecting many other syndromes, for example locomotive syndrome.”

P7 Do you mean muscle strength or muscle mass, please specify

RESPONSE: thank you for comment, since strength training affects both the muscle strength and muscle mass we changed the text as follows: “Much research has been conducted on the effect of strength protocols on muscle mass and muscle strength which incorporate large muscle groups at intensities around 70–80% of 1RM (one repetition maximum), volumes from two to three sets of 10–12 repetition, and rest periods of short to medium duration (60–90 s) [54-55].”

General warm up? What is this based on?

Response: Thank you for the comment. The word “general” was removed from the manuscript. The warm up includes a typical exercise protocol used in practise. (used to specify whole body exercise program).

Reviewer: 1 Reviewer Name: Gustavo Monnerat Institution and Country: Federal university of Rio de Janeiro, Brazil Please state any competing interests or state ‘None declared’: none declared Please leave your comments for the authors below no additional comments

Reviewer: 3 Reviewer Name: Jacqueline Dawson, PhD Institution and Country: California State University, Long Beach, USA Please state any competing interests or state ‘None declared’: None declared. Please leave your comments for the authors below commend the authors on the revised version of the manuscript. Please see the points below on a few items that are still unclear.

Strength training intervention – p. 14, line 27. “Since unilateral exercise (e.g. squats) …” Do you mean to say split squats? A squat is a bilateral exercise.

RESPONSE: Thank you for your comment, the reference mentioned used the term “squats” for one leg squats, which is a unilateral exercise, similar to split squats used in our protocol. For the clarity of the text, we changed the text as follows: “Since unilateral exercises (e.g. one leg squats) develop similar…”.

2) In the sample size section, it is mentioned that fat, lean mass and MVC on leg press were used to determine the sample size. Please clarify which of these variables were actually used to determine the sample size. Also, it would be helpful to readers to state the value of the change (kg) ± SD for the variable that you are using for the calculation. Finally, how are you accounting for attrition?

RESPONSE: Thank you for a comment, we changed the part about sample size calculation, specified the variables as follows:

“The pre-existing data from our previous 12-week exercise intervention study related to fat (5.6% decrease p=0.002) & lean body mass (1.8% increase, p=0.047, DXA) and that of maximal voluntary contraction force measured on linear leg-press (31% increase, p<0.0001, 1RM). Lean body mass was used to determine sample size for the population of the designed intervention study. The Type I error probability was set at 0.05 and the power to 0.90 and 0.95. Results indicate that 22 patients per group will be sufficient to detect exercise intervention related change of 1,06 ± 1,56 kg (average ± SD) of lean body mass at the power of 0.90, accounting for the 10% patients drop-out.”

3) Please add the expected n to the methods section of the abstract.

RESPONSE: Thank you for the comments, we added following text to the abstract: “Lean mass is used to determine sample size indicating, that 22 subjects per group will be sufficient to detect intervention related changes at the power of 0.90”.

4) In the statistical analysis section, a description of how you will manage missing data is absent. Please also list the software that you will use to perform the statistical analysis.

RESPONSE: Thank you for a comment, we added: “All statistics were performed using a statistical software Statistical Package for the Social Sciences (SPSS) 21.0 (IBM Inc., Armonk, New York, U.S.,) and p values < 0.05 will be considered significant. considered significant. Missing Endpoint data will disqualify patient from the endpoint analysis. Missing single value, of training progression records will be replaced by the last observed value.”

5) How is exercise program and study safety assessed? It is mentioned that familiarization and a qualified instructor will oversee the training, but how are you tracking injuries, adverse events, etc. and how does this affect the calculation of compliance vs. adherence?

RESPONSE: thank you for comment, it is an important issue. We added the following text: “All adverse events during the training session will be written down into paper spread sheet and processed afterwards. Each training session will be monitored with an attendance list, with minimum 85% attendance during the study. Each session will be marked as successfully completed when at least 80% from the total volume and intensity of the training protocol planned for the particular training session is performed. “

6) Is dietary intake monitored? Even though you mention that you are not controlling for diet, tracking dietary intake, whether through a log or other means, can certainly strengthen your protocol. As you’ve already begun data collection, this may not be possible, in which case it should be listed as a limitation of your study design.

RESPONSE: Thank you for the comment, the dietary intake is no doubts an important issue. However, as we have already begun with data collection, we listed this as a study limitation according your suggestion as follows: “Another limitation is that the participants will be asked to not change their habitual dietary intake during the intervention, but the actual intake is not monitored. “

VERSION 3 – REVIEW

REVIEWER	Jacqueline Dawson California State University, Long Beach, United States
REVIEW RETURNED	15-Jun-2019

GENERAL COMMENTS	I appreciate the author's commitment to improving their manuscript. This version is much improved from the original and I applaud their efforts. I have no additional comments.