Author Response 1

Comparative measurement properties of constant work rate cycling and the endurance shuttle walking test in COPD: the TORRACTO® clinical trial

Reviewer 1

The analysis done is excellent and purposeful.

> Thank you for your comment.

Please try to correlate results of these tests between them at both baseline and 6 weeks in order to better document their concordance.

> We have added Figure 1D to show endurance time at Week 6 in the two exercise tests, and amended the text on page 11 accordingly:
>
> Distribution and comparison of exercise duration for CWRCE and ESWT at baseline and Week 6
>
> The distributions of the CWRCE and ESWT endurance times at baseline are presented in Figure 1A and B. In both cases, the distribution of exercise duration showed a rightward skew toward long exercise duration. A scatter plot of exercise duration during cycling and walking at baseline and at Week 6 is provided in Figures 1C and 1D.

Reviewer 2

This is a very interesting study in which two different exercise tests have been compared within the frame of the TORRACTO study in patients with COPD. I have a few comments that are summarized below for the authors.

> Thank you for your comment. We hope to address your points below.

1. A flow chart depicting how the patients were selected and/or discarded and finally studied should be included in the manuscript. This will help the reader to better understand how the patient selection was conducted.

> Please see the flowchart (Figure S1), which is now included in the supplement. This shows how patients were selected for the sub-study. For more information regarding participating centres and the numbers of patients from each centre, please see table S1.

2. For the purpose of the current investigation, did patients from all 25 centers participated in the study? Otherwise, which centers contributed with patients in the current investigation?

> The TORRACTO trial included 58 centres in 10 countries; however, only 25 centres with experience of conducting ESWT participated in the ESWT substudy (present investigation). Patients from all 25 centres participated in the sub-study. Please see table S1 for more detail.

3. How many patients per center were included in the TORRACTO trial and in the present investigation? This information should also be included in the flow chart.

> The TORRACTO trial included 58 centres in 10 countries; however, only 25 centres with experience of conducting ESWT participated in the ESWT substudy (present investigation).

We have included a table in the supplement (Table S1) with some detail regarding the participating centres and numbers of patients. However, we feel that including the number of patients in each centre is too detailed and does not contribute to the objectives of this manuscript.

4. If patients were recruited from 25 centers, how were the criteria for patient recruitment and exercise test validity ensured in each center? Who supervised the exercise tests in each center? How was the homogeneity among centers ensured?

We have added the below text on standardisation to page 3 of the supplement:

Test standardization measures

Trial centres were selected to ensure investigators and their staff were qualified to conduct the trial and that the facilities were adequate for trial conduct.

Great care was taken to ensure that all exercise tests were performed according to the highest current published standards on cardiopulmonary exercise testing. Detailed training on the conduct of the exercise tests was provided to site staff at pre-study Investigator Meetings, including a workshop in which all participants observed the conduct of the exercise tests. A clinical research associate (CRA) at each site checked that all site personnel conducting the exercise tests had sufficient qualifications, and that they demonstrated competency in conducting the tests on a healthy volunteer. In addition, a global exercise trainer was contracted by the sponsor to work with selected exercise specialists in each country to ensure consistency in the conduct of the cycle ergometry tests and measurements. Sites were not able to enrol patients until signed copies of all proficiency tests were received from the regional trainer. During the study, the CRAs were responsible for checking the exercise data and identifying any concerns. To supplement the information on exercise testing in the protocol, a Manual of Procedures was developed. Sites were also required to perform a monthly biological quality control test using a healthy volunteer to ensure stability of the equipment and gas analysis systems.

5. Reference values used for each exercise test should be described and cited. How were comparisons of results obtained in each center made in the present study? For the purpose of the current study, in which centers was the ethical approval obtained?

We have added a sentence to the Methods, Study design on page 5, as follows:

In total, 25 centres participated in the ESWT sub-study (Supplementary Figure S1 and Table S1), and they received ethics approval for both the main study and the ESWT sub-study. This secondary analysis of TORRACTO® involved patients in whom both the ESWT and the CWRCE tests were performed.

Regarding cross-centre comparisons, the specific purpose of our analysis was to compare ESWT and CWRCE for the full dataset in a multicentre study design. In addition, due to the limited number of patients and centres included, we are not able to explore differences between centres.

6. The study hypothesis and objectives should be better defined in the abstract and in the last paragraph of the Introduction section.

We have amended both the abstract and introduction to clarify the hypothesis and objective of this sub-analysis (detailed below). Authors: Please let us know if you have any further amends.

Abstract (first paragraph; page 2 main manuscript)

Background: Exercise tolerance is an important endpoint in COPD clinical trials. Little is known about the comparative measurement properties of constant work rate cycle ergometry (CWRCE) and the endurance shuttle walking test (ESWT). The objective of this sub-analysis of the TORRACTO® study was to directly compare the endurance measurement properties of CWRCE and ESWT in patients with COPD in a multicentre, multinational setting. We predicted that both tests would be similarly reliable, but that the ESWT would be more responsive to bronchodilation than CWRCE.

Introduction (last paragraph; page 4)

The TORRACTO® study evaluated the effect of tiotropium/olodaterol (T/O) compared with placebo on endurance during CWRCE in patients with COPD after 12 weeks of treatment.[19] Endurance during the ESWT was evaluated in a subset of patients.[19] The objective of this analysis was to compare, for the first time, the measurement properties of CWRCE and ESWT in patients with COPD in a multicentre, multinational setting at baseline and after 6 weeks of treatment. Specifically, we compared the test–retest reproducibility and the responsiveness of CWRCE and the ESWT, testing the hypotheses that both tests would be similarly reliable, but that the ESWT would be more responsive to bronchodilation than CWRCE. We also compared the relative intensity of breathing discomfort and leg discomfort, as well as the locus of symptom limitation, during these two exercise testing modalities.

Reviewer 3

This study compares the performance and reliability of two methods used to measure endurance time in COPD, as a secondary analysis of the TORRACTO clinical trial. This study sounds pretty important in this field, especially looking to define multicentre guidelines. Despite the importance and convenience of this paper, I have some suggestion and a central concern regards methodology.

Thank you for this summary. We hope to address your points below.

1. The conclusions in the abstract and the main manuscript are not similar, and they should be more concordant a specific.

The conclusions in the abstract and main manuscript were amended (detailed below).

Abstract (last paragraph; page 2-3)

Conclusions: Both exercise tests performed well in a multicentre clinical trial. Although the locus of symptom limitation differed between the two tests, both were reliable and responsive to bronchodilation. For future clinical trials, the choice of test should depend on the study requirements.

Discussion (last paragraph; page 21)

The CWRCE and ESWT both performed well in the context of a multicentre clinical trial. The locus of symptom limitation differed between the two tests; however, both tests were

reliable and responsive to bronchodilation. For future trial design, the choice of test should depend upon the requirements of the study, including the need to collect physiological response data, which can be done more conveniently during stationary cycling.

2. The CWRCE seems to present heteroscedasticity as presented on the Bland-Altman plot (Figure 2d). The authors should describe and calculate this, discussing the findings.

The following statement was added to demonstrate the increase in variability in endurance time over time (Page 11): The reproducibility of ESWT and CWRCE performance was also confirmed with Bland–Altman plots (Figure 2C and D). The variability of endurance time seems to increase over time during both ESWT and CWRCE; mean (SE) endurance time in the T/O arms was 471.49 (28.74) seconds on Day 46 and 472.73 (30.50) seconds on Day 88 during ESWT, and 561.50 (31.16) seconds on Day 43 and 576.69 (36.28) seconds on Day 85 for CWRCE. The reproducibility of both exercise modalities assessed using ICC was 0.75 for the ESWT and 0.56 for CWRCE (Table 2). Although the ICC for CWRCE was numerically smaller, the difference did not reach statistical significance (there was an overlap in ICC 95% CIs between the tests).

3. Could the authors explain why use both the Pearson correlation coefficient and ICC? There is a specific reason for presenting both methods?

The Pearson product–moment correlation was used to evaluate the linear correlation between exercise time during both the CWRCEs and ESWTs. The intraclass correlation coefficient was used to assess performance reproducibility.

4. Considering the importance of establishing narrow limits on endurance time during cycle ergometry between 3-8 min, I think the authors should present a sub-analysis of performance and reliability within these limits. I have serious concerns about the data presented, because looking at the figure 1a. there are too much subjects 480sec. This could bias importantly the results.

Thank you for your suggestion. We did not have enough participants for this analysis, and feel it is outside the scope of the current manuscript. However, we have acknowledged this in the discussion for future work.