Author's response to reviews

Title: Effects of an 18-week exercise programme started early during breast cancer treatment: a randomised controlled trial in daily clinical practice

Authors:

Noemie Travier (ntravier@iconcologia.net)
Miranda J Velthuis (M.Velthuis@iknl.nl)
Charlotte N Steins Bisschop (cnsteinsbisschop@gmail.com)
Bram van den Buijs (B.J.W.vandenbuijs@umcutrecht.nl)
Evelyn M Monninkhof (E.Monninkhof@umcutrecht.nl)
Frank Backx (F.J.G.Backx@umcutrecht.nl)
Maartje Los (m.los@antoniusziekenhuis.nl)
Frans Erdkamp (f.erdkamp@orbisconcern.nl)
Haiko J Bloemendal (h.bloemendal@meandermc.nl)
Carla Rodenhuis (C.C.Rodenhuis@umcutrecht.nl)
Marnix AJ de Roos (m.de.roos@zrt.nl)
Marlies Verhaar (mverhaar@zuwehofpoort.nl)
Daan ten Bokkel Huinink (dtbokkelhuinink@diakhuis.nl)
Elsken van der Wall (E.vanderWall@umcutrecht.nl)
Petra HM Peeters (P.H.M.Peeters@umcutrecht.nl)
Anne M May (A.M.May@umcutrecht.nl)

Version: 2 Date: 10 April 2015

Author's response to reviews: see over
RESPONSE TO THE REVIEWERS: MS: 9464988601577244
Effects of an 18-week exercise programme started early during breast cancer treatment: a randomised controlled trial

We thank the editor and the reviewers for their valuable and detailed comments. We include a point-by-point response to their comments and the changes in the manuscript are highlighted with ‘track changes’.

Editors’ comments:

1. Please submit a completed CONSORT checklist (www.consort-statement.org/consort-statement/checklist) as an additional file

Authors’ answer to comment 1:
We uploaded a completed CONSORT checklist.

2. Please include an explanation of how this study provides an advance over previous related studies in the discussion section.

Authors’ answer to comment 2:
We added the following paragraph to the discussion section:
Compared to previous related studies, the PACT study differed in timing of the intervention (i.e., early in the treatment process) and in location of the intervention (i.e. at the treating hospitals). In previous studies, the intervention was mostly delivered by the same physiotherapist(s) at a well-equipped research center. In daily practice, however, the intervention will be given at different sites with different physiotherapists. The PACT study, therefore, used a more pragmatic design. Although physiotherapists worked according to a standardized protocol, the different locations may have added variability and diluted intervention effects. However, external generalizability is increased.

3. Please provide names of the individual ethics committees that gave approval for your study. If this is a long list, it can be uploaded as an additional file

Authors’ answer to comment 3:
We now added the names of all participating hospitals to the ethics statement in the method section.

4. Authors' contributions: please confirm whether all authors confirmed the final version of the manuscript for submission.

Authors’ answer to comment 4:
Yes, all authors confirmed the final version of the manuscript for submission.
Reviewer: Karen Steindorf

Reviewer's report:
This, in general, well-written manuscript describes a randomized two arm trial in which patients in the early phase of treatment for breast cancer received usual care or a mixed aerobic and muscle strength training interventions. The findings from this well-conducted study are important, particularly with regard to minimization of physical fatigue through exercise therapy.

1) Major Compulsory Revisions
The author must respond to these before a decision on publication can be reached. For example, additional necessary experiments or controls, statistical mistakes, errors in interpretation.

Comment 1:
Title, Background, Methods, and Discussion:
It was not clear enough to the reviewer where the relevant point for the extension “in daily clinical practice” in the title is. Also the point with “Explanatory trials” in the background section is not taken well enough. What are the major differences to other trials? At least most of the cited publications were also embedded in normal clinical routines. Did that impact the use of standard procedures for this study? If so, please discuss possible effects on the internal and external study validity.

Authors’ answer to comment 1:
We agree with the reviewer that the extension “in daily clinical practice” in the title is not very clear. Therefore, we deleted this part of the title. We also deleted the sentence about explanatory trials in the introduction, since we agree that some other trials were also embedded in routine care. For example, the physical exercise program in the RCT of Mutrie et al. was performed at eight community exercise facilities, but this study differed from our study in terms of recruitment timing (i.e. on average 6 month post-diagnosis versus within 6 weeks post-diagnosis) and treatment (72% receiving chemotherapy versus 100% of PACT patients receiving chemotherapy). The RCTs of Steindorf et al, Schmidt et al and Courneya et al were designed to answer efficacy questions, whereas PACT was designed as an effectiveness study. Although physiotherapists worked according to a standardized protocol the different locations may have added variability and diluted intervention effects. However, external generalizability is increased.
The PACT study was designed more pragmatic in that not a single research sport center was used but the exercise program was delivered at seven hospitals, with 2-4 physiotherapists working at each site. This resembles daily practice when the program will be launched on a large scale, thus increasing external generalizability of results. Although a standardized exercise protocol was used, the different locations may have added variability and reduced intervention effects. Our pragmatic design might thus have led to a dilution of the intervention effect rather than having affected internal validity, because it does not provoke confounding, selection or information bias.
We added this issue to the discussion (6th paragraph).

Comment 2:
Results and Discussion:
As the PACT study has been a multi-center trial, some discussion of this issue would be appreciated. Did this approach lead to more heterogeneity compared to other studies, and, if so, with regard to which factors. Does this result in a higher generalizability of results?
Authors’ answer to comment 2:
We believe that a pragmatic multi-center trial design with routine daily conditions resembles real world and thus leads to higher generalizability. In a way, heterogeneity between physiotherapists and hospitals belongs to the “intervention package”.
To account for differences in baseline prognostic factors between patients from different hospitals, we used stratified randomization by hospital.

Comment 3:
Description of the study population in the methods and results sections. Please provide more details on the planned study population. All information should be given at one place in the method section. Exercise started within 6 weeks after diagnosis (Abstract) and patients needed to be “scheduled for chemotherapy”. These are very broad inclusion criteria. Some treatment schemes have a scheduled chemotherapy but it would not happen within the intervention window of this study. Were these patients included? Were there any patients that were not under CTx during the intervention period?

Authors’ answer to comment 3:
In the PACT study patients with different types of treatment were included with the prerequisite that CTx was one of the treatments. Furthermore, we aimed for inclusion as early as possible after diagnosis. Therefore, the PACT study was designed in a way that all patients received in any case chemotherapy during the 18-week exercise program. In The Netherlands, patients usually receive RTx for 3 to 4.5 weeks before CTx if they are at low risk of distant metastases and after CTx if they are at higher risk. By starting the intervention within six weeks post-diagnosis, we made sure that all patients participated in the 18-week exercise program during (part of their) CTx.

As planned, all patients who participated in the PACT study indeed received CTx.

We provide all in- and exclusion criteria in the method section (‘Setting and participants’, page 4+5). All patients needed to have chemotherapy as part of their treatment regime. We added more precise information on the radiotherapy treatment scheme: “In The Netherlands, if indicated, patients usually receive radiotherapy for 3 to 4.5 weeks before chemotherapy if they are at low risk of distant metastases (less than 4 positive lymph nodes), otherwise radiotherapy is scheduled after chemotherapy. By starting the intervention within six weeks post-diagnosis, we made sure that all patients participated in the 18-week exercise program during (part of their) chemotherapy.

With regard to the treatment status at baseline, we added the following information to the results section: “The PACT study, with a duration of 18 weeks and a start within 6 weeks after diagnosis, coincided with chemotherapy treatment in all patients. Seventy-two patients had already started chemotherapy at recruitment; 62 patients had not started chemotherapy, but had started radiotherapy. Seventy patients had not started any treatment yet, but would start with radio- and/or chemotherapy during the intervention period. Neo-adjuvant chemotherapy was still rare, and was used in less than 5% of PACT participants. Treatment status at baseline was balanced between groups (Table 1).”

Comment 4:
Also, the description of the final study population in the results section should be improved. Only very late in the discussion the reader learns that patients were heterogeneous with regard to treatment status at the time of the baseline assessment. Patients were equally distributed over 3 groups at baseline, patients who started CTx, patients who started RTx, and patients who had not started adjuvant therapy. Please include this information into Table 1. For how long are patients treated with RTx before CTx in general? There seems to be
some variation in treatment schemes between countries.

Authors’ answer to comment 4:
Information on the three groups has been added to Table 1 and to the results section (see previous comment).
In the Netherlands, adjuvant radiation therapy, if indicated, is given before adjuvant chemotherapy in patients having a low risk of distant metastases (less than 4 lymph node metastases) while patients at higher risk receive adjuvant chemotherapy first followed by adjuvant radiation therapy. Adjuvant radiation therapy lasts from 3 to 4.5 weeks. Neo-adjuvant chemotherapy is only offered to a minority of early breast cancer patients. In the PACT study, <5 % of the participants received neo-adjuvant chemotherapy.

Comment 5:
The reviewer has some concern with these three groups and feels that this point could be elaborated in more detail. As we know from many studies, CTx has the strongest impact on fatigue, thus the three patient groups are presumably very different with regard to fatigue at baseline. As stratification factor only “radiotherapy yes/no before CTx” was used. However, this means that women with or without CTx were pooled and separated from RTx patients. What was the reasoning for this stratification? The expectation from a clinical perspective would be that ongoing CTx or baseline fatigue would have been better stratification factors, especially as fatigue was the primary outcome.
The authors somehow addressed the issue by performing a sensitivity analyses on CTx before randomization and reported non-significant interactions. But this result might just be explained by a restricted power to test this interaction. This general issue should be discussed as limitation of the study, not only as a strength.

Authors’ answer to comment 5:
We agree that treatment may influence the baseline fatigue. Indeed, general and physical fatigue levels were lower in patients who had not yet started with RTx or Ctx (see the table 1R below). However, because treatment status at baseline is similar between intervention and control groups this did not affect our results.
We used RTx (yes/no) as stratification factor in our randomization procedure to take into account the fact that some participants received both CTx and RTx while others only received CTx. By this, treatment strategies were balanced between groups and had no influence on the outcome. Adjusting our primary analyses for the three groups did not change our results (e.g., analyses on physical fatigue adjusted only for radiotherapy (yes/no): between-group change at 18 weeks: -1.3, 95%CI -2.5;-0.1; with adjustment for the three treatment groups instead: -1.3, 95%CI -2.5; -0.1.).

To acknowledge the fact that some patients already started chemotherapy at recruitment, we performed a sensitivity analyses on CTx before randomization and found that the interaction was not significant. Because the between group estimates observed at week 18 for general and physical fatigue were very similar among the participants who had started CTx at recruitment and among those who had not, we do not think that the absence of interaction could be due to restricted power. However, the fact that outcome measurements were not performed at the same time point with respect to treatment is likely to have added variability to our results and possibly diluted part of the intervention effect.

<table>
<thead>
<tr>
<th>Fatigue level at baseline</th>
<th>No treatment at baseline (n=70)</th>
<th>Started with CTx at baseline (n=72)</th>
<th>Started with RTx at baseline (n=62)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General fatigue</td>
<td>9.6 ± 4.5</td>
<td>10.2 ± 3.9</td>
<td>11.5 ± 4.1</td>
</tr>
</tbody>
</table>
Comment 6:
Please also provide more details on the surgical status of the patients. In setting, 1st paragraph, patients with immediate reconstruction are mentioned. What types of surgery are summarized under this term? Do you cover breast-conserving surgeries with this term?

Authors’ answer to comment 6:
All patients in our study had surgery. We have no information on type of surgery, breast-conserving or mastectomy, we only registered whether mastectomy was performed with immediate reconstruction involving the use of tissue expander. In this case, the inclusion period was 10 weeks post-diagnosis instead of 6 weeks.
We changed the corresponding sentence in the methods section into “The 6-week period was extended to ten weeks if patients had a mastectomy with immediate reconstruction involving the use of tissue expander (n=19).”

Comment 7:
How many patients had neoadjuvant CTx and how many adjuvant CTx? Do you see an issue for interpreting your results?

Authors’ answer to comment 7:
All participants in the PACT study received CTx during the 18-week intervention period. Since less than 5% of the participants received neo-adjuvant CTx, we do not expect any influence of neoadjuvant CTx on our results.

Comment 8:
Please also provide some information on several time windows, for example, time between diagnosis and randomization, time between start of first treatment and randomization etc. Were patients with RTx before CTx included later than the CTx only patients?

Authors’ answer to comment 8:
More than 90% of the participants were included within 6 weeks after histological breast cancer diagnosis. About 10% of the patients had a mastectomy and got a tissue expander placed on their chest wall and they were included within 10 weeks after diagnosis. This information is available in the first paragraph of the method section.
The patients who started CTx before randomization started it on average two weeks before randomization, the patients who started RTx before randomization started CTx on average five to six weeks after randomization, and the patients who had not started adjuvant treatment before randomization started CTx on average 3 weeks after randomization.
The time between diagnosis and randomization was the same for the participants who started CTx before randomization and for those who started RTx before randomization. As expected, the time between diagnosis and randomization was shorter for the participants who did not start any treatment at randomization.
Because we stratified the randomization for radiotherapy (yes/no), the intervention and usual care groups were balanced regarding the number of adjuvant treatments received by the participants (CTx only or CTx and RTx) (Table 1).
Because of randomization, the variable indicating the timing of the intervention with respect to treatment start was also more or less equally distributed across the intervention and usual care groups.
Comment 9:
Control group
The reviewer wonders how surprising the contamination of the control was. In the section “Intervention” it reads that exercise programmes are offered routinely to cancer patients having completed primary treatment. What is the situation in The Netherlands for patients during treatment?

Authors’ answer to comment 9:
Exercise programs after completion of cancer treatment have been offered routinely in the Netherlands for over 10 years. These programs are also reimbursed by most health insurers. However, there are no standard exercise programs that start early during cancer treatment. So, we did not expect the high level of contamination in the control group before the start of the study.

Comment 10:
Abstract (Results) and Statistical Methods:
Fatigue was defined as primary endpoint? Does this refer to general fatigue or to any subscale? Was the analysis at 18-weeks or at 36-weeks planned as primary analyses of the trial or what was the primary hypothesis of the trial?

Authors’ answer to comment 10:
The analysis at 18 week was planned as our primary endpoint. We indicate this at the end of the introduction, and added it in the abstract and when describing sample size calculation. For each subscale, we aimed to be able to detect a between-group difference with a medium effect size assuming a 2-unit (+ SD 4) difference. Because fatigue is a multidimensional construct, we did not specify beforehand a subscale of the fatigue questionnaire as the ultimate primary outcome.

Comment 11:
Abstract, Results, 1st sentence:
Please restructure this sentence to improve readability and correctness. Maybe splitting into two parts and using more different separators within the bracket would help. Please be more specific on the fact that the result for general fatigue was not significant.

Authors’ answer to comment 11:
We restructured the sentence as follows: “Intention-to-treat mixed linear model analyses showed that physical fatigue increased significantly less during cancer treatment in the intervention group compared to control (mean between-group differences at 18 weeks: -1.3, 95%CI -2.5;-0.1; effect size -0.30). Results for general fatigue were comparable but did not reach statistical significance (-1.0, 95%CI -2.1; 0.1; effect size -0.23).”

Comment 12:
Abstract
The presentation of the results on the patients’ fitness is misleading. It is not clear that this statement summarizes results of several measured parameters. The effect sizes do not range from 0.25-0.45 but from 0.02-0.45 with 6 out of 16 parameters reaching statistical significance (see Table 4).
Authors’ answer to comment 12:
We changed the description of the fitness results as follows: At 18 weeks, submaximal cardiorespiratory fitness and several muscle strength tests (leg extension and flexion) were significantly higher in the intervention group compared to control, whereas peak oxygen uptake did not differ between groups.

We agree with the reviewers that for leg muscle strength 4 of 8 tests were significantly different between groups (in favor of the intervention group). These were the tests performed at Cybex velocity settings of 60 degrees per second for flexion and extension of the right and left leg. The tests performed with 180 degrees per seconds were not different between groups. We think that, in the abstract, this is too detailed for the audience of BMC Medicine. Therefore, we now state that several muscle strength tests were significantly higher, but we removed the effect sizes.

Comment 13:
Abstract (Conclusions): The last sentence needs rephrasing. Shortening of the first part but being more precise in the second might help.

Authors’ answer to comment 13:
We split the last sentence into two sentences: “At 36 weeks, these effects were no longer statistically significant. This might have been caused by the control participants’ high physical activity levels during follow-up.”

Comment 14:
Background:
Please check if reference 1 is valid for Western populations and provide a more recent source for Reference 3. Furthermore, the meta-analysis of Brown-F on exercise and cancer-related fatigue should be added as important source.

Authors’ answer to comment 14:
Reference 1 refers to a study based on 3106 US cancer survivors (1544 breast cancer) contacted about a year after diagnosis. Participants were mainly white (86%) and from non-hispanic origin (90%). This reference is, therefore, valid for Western populations.

Reference 3 (Lucia 2003) has been replaced by Bower 2014 (Cancer-related fatigue--mechanisms, risk factors, and treatments). We also adjusted the numbers on fatigue prevalence according to the Bower paper: “Fatigue is reported by up to 30–60% of cancer patients during treatment and many years after treatment up to 25–30% of cancer survivors still report fatigue.”

Brown et al. meta-analysis (now ref 7) has been added to the list of references given at the end of the third sentence of the introduction.

Comment 15:
The placing and the corresponding text on references 9 and 10 raises questions as before the quality of previous studies as well as lacking ITT analyses are mentioned. Please be more precise if these criteria are critical for these two publications too.

Authors’ answer to comment 15:
We restructured the introduction. We now indicated that references 9 and 10 (now 10 and 11 after the addition of ref 7) were not included in the meta-analysis.

Comment 16:
Last sentence: either delete or remove to the methods or discussion section.

Authors’ answer to comment 16:
We moved the sentence to the methods section.

Comment 17:
Intervention:
What was the proportion of time within one training session with regard to the 60 minutes for aerobic and strength training? How standardized was the setting for the aerobic training? What type of exercise was it?

Authors’ answer to comment 17:
The 1-hour exercise classes included a warming-up (5 minutes), aerobic and muscle strength training (25 minutes each) and a cooling down (5 minutes). The aerobic training was performed according to a standardized protocol. Dependent on the patients’ preferences, the training could be done on a cycle ergometer, treadmill or crosstrainer. Intensity of the aerobic training was based on the heart rate at the ventilator threshold as determined during baseline CEPT. The aerobic training included interval training of alternating intensity performed with a heart rate at (3 x 2 min increasing to 2 x 7 min) or below (3 x 4 min decreasing to 1 x 7 min) ventilatory threshold. Heart rate and the Borg scale of perceived exertion were monitored during the aerobic training.
We added this information to the methods section, subheading “Intervention”.

Comment 18:
Sample size:
Why were more patients recruited than planned?

Authors’ answer to comment 18:
As mentioned in the method section, in parallel to the inclusion of patients with breast cancer we included patients with colon cancer. We aimed to include 150 patients per cancer type. Unfortunately, inclusion of colon cancer patients appeared to be difficult, which might be due to either the, on average, older age or the lack of a specialized nurse guiding the patients during the whole treatment phase. These specialized nurses played a crucial role in informing potential participants with breast cancer about the PACT study.
We decided to prolong our inclusion period to increase the number of colon cancer patients in the study. We also continued including patients with breast cancer to ensure continuity of the exercise program and the research procedures (e.g. recruitment in the hospitals). Therefore, we included 204 patients with breast cancer instead of the planned 150.

Comment 19:
Please be more precise for the term “tumor receptor” … and add “status”

Authors’ answer to comment 19:
We added “status” to the term “tumor receptor” and we also defined the variable “triple negative / Her2Neu+, ER+ or PR+ / Her2Neu+, ER- and PR- / Her2Neu-, ER+ or PR+” (method section, subheading “Sample size calculation and statistical analysis”).
Comment 20:
Statistical methods:
In general mixed linear models it is possible to include all collected data even if patients were non-completers. But if non-completers differ systematically from completers results may be biased. As it is reported in the results section that non-completers had higher fatigue, it would be interesting if results changed if standard techniques for handling of missing data were applied (complete-case, multiple imputation)?

Authors' answer to comment 20:
Between-group effects were modelled using the outcome measurements at 18 and/or 36 weeks: participants with only baseline data were not included in this analysis. Within-group changes were modelled using outcome measurements obtained at the three time points (i.e. at baseline, and at 18 and/or 36 weeks). Individuals with one or two measurement missing are included in these analyses. We now also provide this information in the 2nd paragraph of the “Sample size calculation and statistical analysis” section.

Including only those participants with complete data (i.e. at baseline, 18 weeks and 36 weeks), we observe larger effects on general and physical fatigue at 18 weeks (see table 2R). Both were statistically significant.

**TABLE 2R.** Complete case analysis (i.e., participants with fatigue assessed at recruitment, 18 weeks and 36 weeks)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Between-group difference (18 week)</th>
<th>Between-group difference (36 week)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>beta 95%CI</td>
<td>beta 95%CI</td>
</tr>
<tr>
<td>General fatigue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usual care</td>
<td>-1.39 [-2.57; -0.21]</td>
<td>-1.07 [-2.23; 0.08]</td>
</tr>
<tr>
<td>Intervention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical fatigue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usual care</td>
<td>-1.67 [-2.96; -0.39]</td>
<td>-0.67 [-1.93; 0.59]</td>
</tr>
<tr>
<td>Intervention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental fatigue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usual care</td>
<td>-0.21 [-1.49; 1.08]</td>
<td>-0.70 [-1.96; 0.56]</td>
</tr>
<tr>
<td>Intervention</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Between-group effects were assessed using mixed models including the measurements obtained at 18 and 36 weeks, adjusted for age, hospital, radiotherapy, use of tissue expander, tumour receptor status and the value of the outcome variable at baseline.

We did not perform multiple imputation (MI) of missings as previous studies have showed that MI performed before using linear mixed models does not increase precision in the estimated rate of change in the end point (Multiple imputation of missing repeated outcome measurements did not add to linear mixed-effects models. Peters SA et. Al. J Clin Epidemiol. 2012 Jun;65(6):686-95).

Comment 21
What happens if you adjust for a variable that mirrors the three pre-treatment groups? Is there a difference between adjustment and testing interaction effects in your model?

Authors’ answer to comment 21:
We repeated the primary analysis adjusting for the 3-category variable previously defined that mirrors the timing of treatment start with respect to recruitment and found no difference with the results of the analyses as presented in the manuscript (see table 3R). This analysis, along with the interaction analysis, indicate that the timing of the start of the intervention, with respect to treatment, does not affect our results.

**TABLE 3R. Primary analyses adjusted for adjuvant treatment at baseline (3 CTx, RTx, none).**

<table>
<thead>
<tr>
<th></th>
<th>Between-group difference (18 week)</th>
<th>Between-group difference (36 week)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment</td>
<td>beta CI</td>
</tr>
<tr>
<td>General fatigue</td>
<td>Usual care</td>
<td>reference</td>
</tr>
<tr>
<td></td>
<td>intervention</td>
<td>-1.02 [-2.15; 0.10]</td>
</tr>
<tr>
<td>Physical fatigue</td>
<td>Usual care</td>
<td>reference</td>
</tr>
<tr>
<td></td>
<td>intervention</td>
<td>-1.31 [-2.52; -0.09]</td>
</tr>
<tr>
<td>Mental fatigue</td>
<td>Usual care</td>
<td>reference</td>
</tr>
<tr>
<td></td>
<td>intervention</td>
<td>0.02 [-1.18; 1.21]</td>
</tr>
</tbody>
</table>

Between-group effects were assessed using mixed models including the measurements obtained at 18 and 36 weeks, adjusted for age, hospital, baseline adjuvant treatment status, use of tissue expander, tumour receptor status and the value of the outcome variable at baseline.

Comment 22:
**Results**
Participants + Figure 1:
- What does “completed the trial” exactly means.

Authors’ answer to comment 22:
The participants who completed the trial are those who did not drop out during the 36 weeks that lasted the study. All of them came for outcome assessment at 36 weeks. This clarification has been added to the results section of the manuscript.

Comment 23:
- 2nd paragraph, 1st line: Not 13 but 25 participants from the control group were lost according to Figure 1. Were the proportions of missing data different for the different variables?

Authors’ answer to comment 23:
We replaced 13 by 25 participants.
The numbers given in Figure 1 correspond to the participants who completed the questionnaires (i.e. assessment of primary outcome fatigue) at weeks 18 and 36 or at week 36. Some patients filled in the questionnaires but did not visit the research center for the physical fitness tests. We added the corresponding numbers for physical fitness outcomes below Table 4 in the manuscript.
The last sentence of the footnote below table 4 now reads:
“Baseline results and within-group differences were based on participants having at least baseline measurements: body weight: 102 Intervention (I) and 102 usual care (UC), aerobic capacity: 101 (I) and 98 (UC), leg strength: 78 (I) and 79 (UC); hand grip: 98 (I) and 100 (UC).

Between-group differences were based on participants for whom measurements at 18 or 36 weeks were available: 90 (I) and 79 (UC) for body weight, 88 (I) and 76 (UC) for aerobic capacity, 69 (I) and 64 (UC) for leg strength, and 90 (I) and 79 (UC) for hand grip”.

Comment 24:
**Fatigue:**
Please rephrase the sentence starting in line 5 and do not use “all” participants if you refer to
mean values for both groups.

Authors' answer to comment 24:
We deleted the "all" and rephrased the sentence as follows “Over the same period, participants of both groups rated, on average to a comparable extent, their fatigue as more frustrating and exhausting.”

Comment 25:
Is the second last sentence in agreement with table 2?

Authors' answer to comment 25:
We assume that the reviewer refers to the following sentence: “At 36 weeks, women reported fatigue levels and fatigue-related feelings that were in general similar to those observed at baseline and similar between groups.”
We believe that this sentence is in agreement with table 2 because in both groups fatigue levels reported at 36 weeks did not differ significantly from fatigue levels reported at baseline (within-group changes) with the exception of mental fatigue in the usual care group which was still higher. Also, we found no significant differences between the exercise and the usual care group.
We rephrased the sentence to be clearer:
“At 36 weeks, women reported fatigue levels and fatigue-related feelings that were similar to those observed at baseline, with the exception of mental fatigue in the usual care group which was reported to be still higher (1.0, 95%CI 0.1;1.9). No significant differences between the exercise and the usual care group were found.”

Comment 26:
Tables 2-4 and S1:
• The column headings “18 weeks post intervention” are misleading. The assessment is 18 weeks after baseline and post intervention but not 18 weeks away from the intervention period.

Authors’ answer to comment 26:
We changed this accordingly in Table 2-4 and S1.

Comment 27:
• In the footnote it reads that the within-group differences in the table are based on 102 participants in each group that had baseline values. However, the within-group differences refer to changes from baseline to the other timepoints where some of these participants did not have data. How could you include all 204 participants?

Authors’ answer to comment 27:
Mixed linear models include all patients having at least one measurement of the outcome variable at one of the time points taken into account by the model.
Within-group changes model fatigue levels taking into account the three time points (baseline, 18 weeks, and 36 weeks), therefore, subjects who only provided baseline measurement are included in this analysis. These analyses included 204 patients for data from questionnaires, and for muscle strength and aerobic capacity slightly different numbers were included (see comment 22 and updated footnote of table 4).
Between-group differences include two time points as the outcome (18 weeks and 36 weeks, baseline was used for adjustment), and these analyses were based on patients with at least one of the two follow-up time points.
Repeating the analyses excluding patients having only baseline measurements yielded to
very similar within-group changes.

Comment 28:
Table 4:
This table needs major changes to increase readability. "Ext 60d" is too technical and will not help many readers from BMC medicine. Muscle strength should read "Leg muscle strength".

Authors' answer to comment 28:
We hope to have been able to increase readability by adjusting Table 4. We clarified that knee flexion and extension strength were measured with an angular velocity of 60 degrees per second.

Comment 29:
Adherence to the study protocol:
Second last sentence: This may be an explanation. However, if this is important it needs to be elaborated in more detail along Table S2. Maybe, this issue would be better placed in the discussion.

Authors' answer to comment 29:
We now present the per protocol results in more detail in the last paragraph of the results section. We also added the following sentence to the discussion: "The present study also has limitations. The high level of physical activity reported by 56% of the controls at 18 weeks may have led to an underestimation of the true effect."
However, we did not want to stress the per protocol results more because those should be interpreted with caution because of selective non-compliance.

Comment 30:
Discussion:
4th paragraph: The BEATE did not exclude patients with baseline depression in general. There are tables where depressed patients were excluded but these were only secondary analyses.

Authors' answer to comment 30:
We agree with the reviewer that in the BEATE study patients with baseline depression were only excluded from specific analyses such as the ones showing the results in the manuscript of patient reported outcomes pre/post-intervention, including quality of life (EORTC QLQ30), depression (CED-D) and cognitive function (trail making test).
In order to take this into account, we replaced the sentence "These differences might be partly explained by the higher baseline scores observed in the present study or the exclusion of patients with baseline depression in the BEATE study."
by
"These differences might be partly explained by the higher baseline scores observed in the present study or by the exclusion of patients with baseline depression in the corresponding quality of life analyses in the BEATE study."

Comment 31:
As some of your participants received not only chemotherapy but also radiotherapy your results should also be set into the context of those studies (e.g. BEST study). As in both settings exercise effects have recently been reported this may help discussing the effect of the three pre-treatment groups.
Authors’ answer to comment 31:
In the PACT study all participants received chemotherapy, whereas radiotherapy was not given in 31% of the participants. For this reason, we decided to focus the comparison of our results on previous studies having assessed the role of exercise during adjuvant treatment only when this treatment involved chemotherapy. In our discussion on effects on physical fatigue, we include both, the BEATE and the BEST studies.

Comment 32:
Discussion, 7th paragraph:
The presented data on the study population with regard to previous treatment needs to be given earlier. The reviewer does not agree with this argumentation. There are also many disadvantages due to this mixture that needs to be discussed (see above).

Authors’ answer to comment 32:
We now included the treatment status at baseline in Table 1 and a description in the results section. In the results section, we report that the baseline characteristics are well-balanced, including treatment status.
Since the baseline treatment status is now reported in Table 1, we deleted the detailed information in the discussion.
We believe that since the three treatment groups were balanced between intervention and control, by design, this mixture is unlikely to have significantly affected the results of our analysis. To acknowledge possible residual confounding, we performed the interaction analysis which did not show any differential effect.

Comment 33:
2) Minor Essential Revisions
The author can be trusted to make these. For example, missing labels on figures, the wrong use of a term, spelling mistakes.
Please delete “also” in the Abstract, Results, 4th line
Please delete “also” in the Background, 2nd last sentence
Intervention, line 6: at least one “,” is missing.
Table 1
Please correct the unit of the BMI.
Please be more congruent with capitalizing letters.
Table 3:
2nd block in the table as a wrong text (18# 36)

Authors’ answer to comment 33:
We changed the Minor Essential Revisions accordingly.

Comment 34:
3) Discretionary Revisions
These are recommendations for improvement which the author can choose to ignore. For example clarifications, data that would be useful but not essential.
You may want to check on the frequency and necessity of the term “also” throughout the manuscript.
Discussion, 5th paragraph, last line: I suggest to replace “can” by “may”.
Acknowledgements: N travier # capitalize T

Authors’ answer to comment 34:
We changed the Discretionary Revisions accordingly.
Reviewer: Siobhan Phillips

Reviewer's report:

Overview
This study describes an 18 week physical activity intervention that was implemented in clinical practice. The intervention was associated with a lower increase in fatigue and increased fitness and muscle strength in compared to usual care during active treatment for breast cancer. Overall, the study is well-written, but could be strengthened by providing more details regarding the intervention, specifying the clinical significance of these findings and providing some additional thoughts on future directions.

MINOR REVISIONS

Abstract

Comment 1:
Please add a statement to the conclusions that discusses the implications of the findings from this study.

Authors' answer to comment 1:
We added the following sentence to the conclusions in the abstract: “A supervised 18-week exercise programme, offered early in routine care during adjuvant breast cancer treatment, showed positive effects on physical fatigue, submaximal cardiorespiratory fitness and muscle strength. Exercise early during treatment of breast cancer can be recommended.”
The conclusion in the manuscript was amended as follows: “Exercise is beneficial during adjuvant breast cancer treatment by reducing the development of fatigue.”

Background

Comment 2:
Please rephrase the statistics on fatigue to state that up to 70% and up to 30% as the estimates of prevalence vary widely in this population.

Authors' answer to comment 2:
As recommended by the other reviewer, we now use a more recent reference for the fatigue prevalence (Bower et al 2014). The corresponding sentence is adapted as follows: “Fatigue is reported by up to 30% to 60% of cancer patients during treatment and up to 25% to 30% still report fatigue many years after treatment.

Comment 3:
Sentence starting with “The intervention…” I think you are missing the word “and” between diagnosis and was.

Authors' answer to comment 3:
As suggested by the reviewer we rephrased this sentence into “The intervention started as early as possible after breast cancer diagnosis and was offered at the patients’ treating hospital.”

Methods
Setting and Participants
Comment 4:
What is meant by stage M0? Please clarify.

Authors’ answer to comment 4:
M0 means no distant metastasis. We have added a more detailed description of the eligible cases in the methods section.

Comment 5
Were there any inclusionary/exclusionary criteria relevant to current physical activity levels? Please comment on this within the text.

Authors’ answer to comment 5:
There were no in- or exclusion criteria used for current physical activity levels. We added the following sentence to the first paragraph of the method section: “Inclusion was irrespective of the patients’ current physical activity level.”

Comment 6
How exactly were women recruited? It seems that the sample may be a bit healthier and more active than would be expected. Could this have something to do with how potential participants were identified?

Authors’ answer to comment 6:
Women were recruited by the oncology nurses or treating medical specialists in the hospitals during a regular outpatient clinic visit. We asked the nurses and the medical specialists to invite every breast cancer patients that meets the inclusion criteria to participate in the PACT-study. However, it might well be that the nurses and medical specialists invited, probably unconsciously, the healthier and more active women.

Comment 7
The last sentence is worded somewhat awkwardly.

Authors’ answer to comment 7:
We rephrased the sentence into: “Colon cancer patients were also included in the PACT study. Results for colon cancer patients will be presented separately to be able to address site-specific issues.”

Comment 8:
Did recruitment vary across sites? Were participants similar?

Authors’ answer to comment 8:
Recruitment numbers varied across centers, for example, due to different total number of eligible patients per center (e.g. academic versus non-academic). In order to take this variability into account and also the possible difference in prognosis of participants in the different hospitals at baseline, randomization was performed within each center.

Intervention

Comment 9
Please add “social” in between Bandura’s and cognitive.
Authors’ answer to comment 9:
We added “social” in between Bandura’s and cognitive.

Comment 10
How exactly were the preference and fitness levels determined to individualize the training?

Authors’ answer to comment 10:
Preferences were determined during intake by the physiotherapist prior to start of the exercise program. Information about the patient’s sports preferences, regular sport activities, requirements at home and work were inventoried. This information was used to design a personalized aerobic training. This included for example the equipment choice (cycle ergometer, treadmill or cross-trainer) and the advice on type and duration of activity for three other days (in addition to the two supervised sessions per week).

Fitness levels were determined by means of a cardiorespiratory exercise test (heart rate at the ventilator threshold) and 1-repetition maximum muscle strength tests in order to individualize the aerobic and the muscle strength training, respectively.

We added the following information to the method section, subheading “Intervention”: “The exercise training was individualized to the patients’ preferences inventoried during the first exercise session and fitness level assessed by means of a cardiopulmonary exercise test and 1-repetition maximum (1-RM) muscle strength tests.”

Comment 11
Please specify the length of the total interval session and each individual intervals and what is meant by “low ventilatory threshold”. Was this monitored during exercise?

Authors’ answer to comment 11:
We added detailed information about the length of interval sessions in the method section, subheading “Intervention”: “Intensity of the aerobic training was based on the heart rate at the ventilatory threshold as determined during baseline CPET. The aerobic training included interval training of alternating intensity performed with a heart rate at (3 x 2 min increasing to 2 x 7 min) or below (3 x 4 min decreasing to 1 x 7 min) ventilatory threshold. Heart rate and the Borg scale of perceived exertion were monitored throughout the aerobic training.”

Comment 12:
How was training intensity re-evaluated every 4 weeks?

Authors’ answer to comment 12:
Training intensity was re-evaluated every four weeks by a submaximal cardiopulmonary exercise test (to evaluate heart rate at the ventilatory threshold) and by repeating the 1-RM muscle strength tests. We added this information to the method section, subheading “Intervention”.

Comment 13:
Was the activity prescribed outside of the program aerobic activity? How was adherence to this portion of the intervention monitored?

Authors’ answer to comment 13:
The activity prescribed outside of the program was indeed aerobic activity. We added this information in the following sentence (Methods section, subheading “Intervention”):
“This should include an aerobic component of moderate intensity in agreement with the participants’ fitness and desires.”

The adherence to the aerobic activity outside of the program was recorded by the patient in an exercise log. We also added this information: “Adherence to the exercise recommendation was recorded in an exercise log.”

We now also provide information on the adherence to the activity prescribed outside of the program (i.e. adherence to the Dutch Guideline of Physical Activity) in the result section, subheading “Adherence”: “Patients reported to be physically active according to the Dutch Guideline of Physical Activity in 11 (interquartile range 6-14) of the 18 weeks.”

Comment 14
What is meant by “exercise programmes offered routinely”? Please clarify.

Authors’ answer to comment 14:
Exercise programs after completion of primary cancer treatment have been offered in The Netherlands for over 10 years and are reimbursed by most health insurers. Therefore, these programs were considered as usual care.

We rephrased the paragraph to be clearer: “Participants randomised to control received usual care and were asked to maintain their habitual physical activity pattern up to week 18. Then they were allowed, for ethical reasons, to participate in exercise programmes, which have been offered in the Netherlands to cancer patients after completion of primary treatment for over 10 years and are thus part of usual care.”

Sample Size and Statistical Analysis

Comment 15
In the statement starting with “With the current number of participants…,” please specify the n.

Authors’ answer to comment 15:
We have specified the n (n = 204) in the corresponding sentence.

Comment 16:
How was adherence to the 210 minutes of exercise assessed?

Authors’ answer to comment 16:
Adherence to the 210 minutes of exercise was assessed with the SQUASH physical activity questionnaire.
We now indicate this in the methods section, subheading “Sample size calculation and statistical analysis”.

Comment 17:
Did the authors consider a sub-analyses examining those who were adherent in the intervention group compared to those who did meet or exceed the PA recommendations for the control group to provide greater insight into what the potential effects would be in comparison to those who were not doing the same amount of exercise?

Authors’ answer to comment 17:
We performed a per protocol analysis among adherent participants. In this analysis, we excluded women from the intervention group with levels below 210 minutes and from the control group with levels above 210 minutes of moderate-to-vigorous physical activity per week assessed with the SQUASH questionnaire. Results are shown in supplemental Table S2. We now add more information of this analysis in the result section, subheading “Per protocol analyses”:

“Per-protocol analyses showed for both general and physical fatigue, moderate significant differences between participants in the exercise and the usual care group who adhered to the protocol in favour of the intervention group with effect sizes of -0.54 and -0.77, respectively (supplemental table S2).”

Outcome Measures

Comment 18
Was aerobic capacity based on volitional termination of the exercise test?

Authors’ answer to comment 18:
The cardiopulmonary exercise test was terminated at the basis of the patients’ symptoms or at the physicians’ discretion. We added this information to the manuscript (Method section, subheading “Outcome measures”).

Comment 19:
Please provide information on the reliability/validity of the PA measure used.

Authors’ answer to comment 19:
The SQUASH questionnaire was validated in the Dutch population (reference (also in our manuscript): Wendel-Vos GC et al. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003; 56:1163-9.). Results showed that the overall reproducibility of the SQUASH was r=0.58 (95%-CI 0.36–0.74)) and the reproducibility of the separate questions varied between 0.44-0.96. The correlation between the Computer Science and Applications (CSA) Activity Monitor readings and the total activity score was 0.45 (95%-CI 0.17–0.66).

Results

Comment 20:
Were all participants lost to follow-up excluded from analyses? If yes, please specify.

Authors’ answer to comment 20:
No, we did not exclude all participants lost to follow-up. Mixed linear models for between-groups changes included all patients with at least one of the two follow-up measurement of the outcome (at 18 weeks or 36 weeks, baseline was adjusted for) and thus included patients with incomplete data. Within-group changes included all three measurements as the outcome (baseline, 18 weeks and 36 weeks). All patients with at least one measurement were included. For different outcomes different numbers were used. This information can be found in footnotes of the tables.

Comment 21:
Please present data on adherence to the intervention protocol prior to the other results.
Authors’ answer to comment 21:
As suggested by the reviewer, we now present the data on adherence prior to the other results.

Comment 22:
Did the intervention participants engage in significantly more activity post-intervention than pre-intervention? Did they engage in more activity post-intervention/at follow-up than control participants?

Authors’ answer to comment 22:
At 18 weeks, as expected due to chemotherapy, both the intervention and the usual care groups had decreased their level of physical activity. The decrease was higher among the participants of the usual care group, however, the difference was not statistically significant. At 36 weeks both groups still showed decreased level of physical activity when compared to baseline, and there were no between-group differences. This decrease was similar in both groups.

Discussion

Comment 23
Why is physical fatigue most likely to be influenced by exercise? Please explain.

Authors’ answer to comment 23:
We only speculated that physical fatigue might be the fatigue subscale which is most likely to be influenced by physical exercise because exercise directly changes physical fitness. We now formulated the sentence less strongly: “Physical fatigue might be the fatigue dimension most sensitive to exercise”.

Comment 24:
Mechanisms paragraph: What about psychosocial mechanisms? There are some studies which suggest PA may decrease fatigue through its effects on self-efficacy and depression.

Authors’ answer to comment 24:
We agree with the reviewer that psychosocial factors can mediate effects of exercise on fatigue as well. We added information about the psychosocial mechanism from Buffart et al. (Buffart et al. Mediators of physical exercise for improvement in cancer survivors' quality of life. Psychooncology. 2013;10) to the corresponding paragraph in the discussion: “Also, psychosocial mechanisms might play a role. Buffart et al. showed that a supervised exercise program resulted in increased physical activity, general self-efficacy and mastery in patients with cancer after treatment, which led to reduced fatigue and distress and consequently improved QoL.”

MAJOR REVISIONS

Methods
Intervention

Comment 25:
Given that this intervention was implemented in practice, I would have liked more information
on the flow of participants through the intervention (from recruitment to enrollment) to understand how participants were identified and who was involved (from a staffing point) in the trial.

Authors’ answer to comment 25:
Women were recruited by the oncology nurses or medical specialists who were treating them in the different hospitals during a regular outpatient clinic visit. This information is given in the first paragraph of the Methods section. Nurses and medical specialists were asked to invite all breast cancer patients who met the inclusion criteria of the PACT-study. As indicated in Figure 1, 426 women meeting the inclusion criteria were actually approached, and 52% (222/426) of them refused to participate for various reasons listed in Figure 1 and mentioned in the first paragraph of the results section.

Comment 26:
How was fidelity to the intervention measured across sites?

Authors’ answer to comment 26:
At all study sites, the attendance rate for the exercise sessions and the compliance with the protocol were recorded in each patient’s Case Record Form. We added this information to the method section, subheading “Adherence”.

Results

Comment 27
Aside from statistical significance, were the differences observed clinically meaningful even if they were/were not statistically meaningful?

Authors’ answer to comment 27:
We provide information on effect sizes to acknowledge clinically relevance. We calculated Cohen’s effect sizes: According to Cohen, effect sizes < 0.2 indicate ‘no difference’, effect sizes of 0.2 to 0.5 indicate ‘small differences’, effect sizes of 0.5 to 0.8 indicate ‘moderate differences’ and effect sizes ≥ 0.8 indicate ‘considerable differences’. At 18 weeks, we find small beneficial effects of the exercise intervention for e.g. physical fatigue (ES 0.30), VO$_2$ at ventilatory threshold (ES 0.30) or several muscle strength tests (highest ES 0.45).
For some variables (e.g. general fatigue and role functioning), we found small albeit non-significant effects. Given the large number of tests we performed, we did not want to over-emphasize these non-significant effects.

Discussion

Comment 28:
I think a major limitation of this study is the fact that it seems as though the participants included may have been the ones that needed the program given their high pre-diagnosis activity levels.

Authors’ answer to comment 28:
We agree with the reviewer that physical activity levels of our study participants were generally high. We added the following sentence to the limitations in the discussion: “Participants in the current study reported on average a high pre-diagnosis physical activity
level and might thus not be the ones who needed the program most."

Comment 29:
Some other potential weaknesses of this study are: a) the lack of an objective PA measure so all findings are based on self-report and b) the use of a submax exercise test which may be less sensitive to change, especially in women who are already pretty active and likely relatively fit. Additionally, it is hard to tell what type of activity (since it is combined aerobic and resistance training) might be driving some of the results.

Authors’ answer to comment 29:
a) We agree with the author that the lack of an objective PA measure is a limitation of this study and we added this limitation to the discussion section.

b) We actually used a maximum exercise test in the PACT study (see method section “Aerobic capacity”). From the cardiopulmonary exercise test, we retrieved peak oxygen uptake and peak power output (at exhaustion) as well as oxygen uptake and power output at the ventilatory threshold. We found beneficial effects at the submaximal level (at ventilatory threshold) but not the maximum level (at exhaustion).

c) We agree with the reviewer that it is hard to tell what type of activity (since it is combined aerobic and strength training) might be driving some of the results and we also added this point to the limitations (“We offered a combined aerobic and strength exercise program to the patients. Therefore, we cannot distinguish what type of activity might have driven our results.”). We had chosen a combined aerobic and strength program because both components might be important for preventing fatigue and according to general exercise principles a combination is recommended.

Comment 30:
Please provide some future directions in the context of this study.

Authors’ answer to comment 30:
We added the following future directions to the article:
“This study shows that exercise during adjuvant treatment of breast cancer is beneficial in reducing fatigue. Women with low physical activity levels might benefit more but seem less interested in exercise programs. Future studies should elucidate patients’ attitude, motivation and barriers towards participation in exercise programs in order to specifically design exercise programs for the less active patients.”