



# Effects of exercise in breast cancer patients: implications of the trials within cohorts (TwiCs) design in the UMBRELLA Fit trial

Roxanne Gal<sup>1</sup> · Evelyn M. Monninkhof<sup>1</sup> · Carla H. van Gils<sup>1</sup> · Rolf H. H. Groenwold<sup>2</sup> · Sjoerd G. Elias<sup>1</sup> · Desirée H. J. G. van den Bongard<sup>3</sup> · Petra H. M. Peeters<sup>1</sup> · Helena M. Verkooijen<sup>4</sup> · Anne M. May<sup>1</sup>

Received: 8 April 2021 / Accepted: 8 August 2021 / Published online: 24 August 2021  
© The Author(s) 2021

## Abstract

**Purpose** The Trials within Cohorts (TwiCs) design aims to overcome problems faced in conventional RCTs. We evaluated the TwiCs design when estimating the effect of exercise on quality of life (QoL) and fatigue in inactive breast cancer survivors.

**Methods** UMBRELLA Fit was conducted within the prospective UMBRELLA breast cancer cohort. Patients provided consent for future randomization at cohort entry. We randomized inactive patients 12–18 months after cohort enrollment. The intervention group ( $n=130$ ) was offered a 12-week supervised exercise intervention. The control group ( $n=130$ ) was not informed and received usual care. Six-month exercise effects on QoL and fatigue as measured in the cohort were analyzed with intention-to-treat (ITT), instrumental variable (IV), and propensity scores (PS) analyses.

**Results** Fifty-two percent ( $n=68$ ) of inactive patients accepted the intervention. Physical activity increased in patients in the intervention group, but not in the control group. We found no benefit of exercise for dimensions of QoL (ITT difference global QoL: 0.8, 95% CI = - 2.2; 3.8) and fatigue, except for a small beneficial effect on physical fatigue (ITT difference: - 1.1, 95% CI = - 1.8; - 0.3; IV: - 1.9, 95% CI = - 3.3; - 0.5, PS: - 1.2, 95% CI = - 2.3; - 0.2).

**Conclusion** TwiCs gave insight into exercise intervention acceptance: about half of inactive breast cancer survivors accepted the offer and increased physical activity levels. The offer resulted in no improvement on QoL, and a small beneficial effect on physical fatigue.

**Trial registration** Netherlands Trial Register (NTR5482/NL.52062.041.15), date of registration: December 07, 2015.

**Keywords** Trials within cohorts · TwiCs · Breast cancer · Physical activity · Quality of life · Fatigue

## Introduction

Fatigue is the most often reported side-effect of breast cancer and its treatment, which can persist for years and negatively affect quality of life (QoL) [1–3]. Meta-analyses of randomized controlled trials (RCTs) have shown that

physical exercise has positive effects on fatigue and QoL in patients with cancer [4–8]. However, these effects were often small, which may be partly due to the inability of blinding the intervention [9]. Patients who decided to participate are generally motivated to exercise and may be disappointed when allocated to the control group. Consequently, they drop-out or start exercising by themselves, the latter resulting in contamination and dilution of the intervention effect [10]. Other disadvantages of conventional RCTs are time-consuming accrual and inclusion of a selective study sample [11, 12].

The Trials within Cohorts (TwiCs) design, also known as the cohort multiple randomized controlled trial (cmRCT) design, was proposed as an alternative to conventional pragmatic RCTs, and has the potential to overcome above mentioned challenges [13–15]. Using this design, the intervention study is performed within a prospective cohort. Compared to conventional RCTs, the TwiCs design can

✉ Anne M. May  
A.M.May@umcutrecht.nl

<sup>1</sup> Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, STR 6.131, PO Box 85500, 3508 GA Utrecht, the Netherlands

<sup>2</sup> Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

<sup>3</sup> Department of Radiation Oncology, Amsterdam University Medical Centers, Amsterdam, The Netherlands

<sup>4</sup> Division of Imaging and Oncology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

lead to more efficient patient recruitment, generalizability of results may be improved, and TwiCs allows evaluation of patients’ acceptability of the intervention [14–17].

The current UMBRELLA Fit study is the first trial applying the TwiCs design in the field of exercise-oncology. UMBRELLA Fit examined the effect of a 12-week supervised exercise intervention on QoL and fatigue. Moreover, we aimed to evaluate the applicability of the TwiCs design in the field of exercise-oncology and the implications of the TwiCs design on effect estimation and interpretation of the results.

## Materials and methods

### Study design and procedures

The study was approved by the Ethics Committee of the University Medical Center Utrecht (UMCU). The UMBRELLA Fit study is a pragmatic, two-arm RCT using the TwiCs design and is conducted within the ‘Utrecht cohort for Multiple BREast cancer intervention studies and Long-term evaLuAtion’ (UMBRELLA) [13, 18, 19]. Since September 2013, all patients with breast cancer who are referred for radiation treatment to the UMCU were invited to participate in the UMBRELLA cohort. At enrollment, first stage informed consent is asked for collection of clinical data and patient reported outcome (Fig. 1) [20]. Additionally, patients could provide broad consent for randomization to future intervention studies. When allocated to an intervention arm (in the future), they will be offered the intervention and asked for (second stage) informed consent when accepting the intervention. Patients allocated to the control arm were not notified about the study and their cohort data were used to estimate intervention effectiveness.

UMBRELLA Fit recruited from October 2015 to March 2018. Women participating in the UMBRELLA cohort meeting the UMBRELLA Fit inclusion criteria (see below) were randomly allocated to either intervention or control with a 1:1 ratio, stratified by time since cohort enrollment (12 or 18 months). Randomization was performed by an independent data manager, using a computer-generated randomization list. Patients allocated to the exercise intervention group received an offer by mail to participate and were contacted by telephone to further explain the study. Patients who refused the offer received usual care. When patients refused because of bad timing of the offer, they got the option to start the exercise program at a later stage. Patients allocated to the control group were not informed about the study, and received usual care. After study completion, all patients in the UMBRELLA cohort, irrespective of participation in this study, were informed by a newsletter.

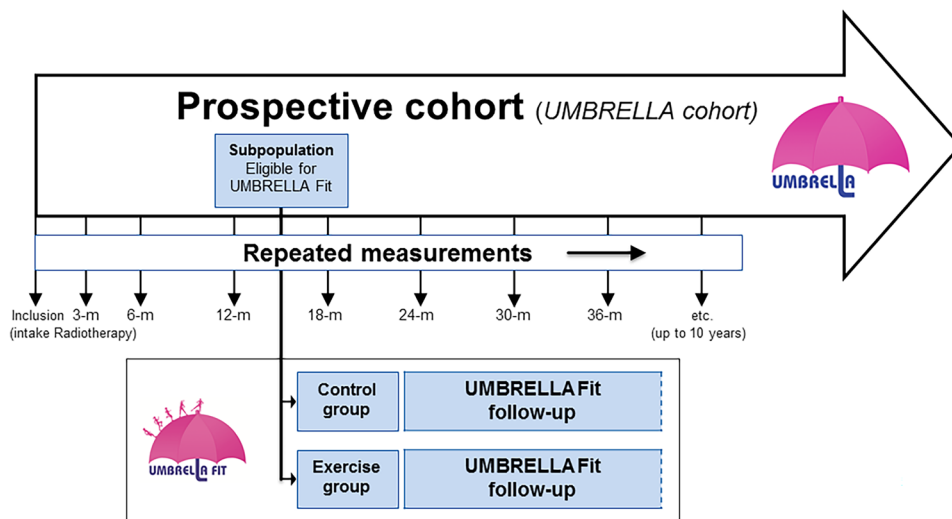
### Participants

Patients eligible for the UMBRELLA cohort [19], were eligible for UMBRELLA Fit when meeting the following inclusion criteria: (1) broad informed consent for randomization to future intervention studies; (2) 18–75 years of age; (3) completion of the 12- or 18-month cohort questionnaire; (4) cancer treatment completed (except for hormonal treatment); and (5) a physically inactive lifestyle (< 150 min/week performing moderate-to-vigorous leisure time and sports activities).

### Intervention

The 12-week supervised exercise intervention comprised two weekly 1-h combined aerobic and resistance training sessions at a physiotherapist center close to the patient’s

**Fig. 1** The UMBRELLA Fit trials within cohorts (TwiCs) design ( adapted from Relton et al.)



home (see details in the supplement). Patients were also encouraged to be physically active with moderate intensity for at least 30 min all days [21].

### Outcome measures

Outcome measures were obtained from routine UMBRELLA cohort measurements. The questionnaire that was completed either 12 or 18 months after cohort enrollment served as baseline and the cohort questionnaire 6 months later was used as follow-up measurement.

### Quality of life

Quality of life (primary outcome) was measured using the Global QoL score and the five functional scales of the EORTC QLQ-C30 [22]. A QLQ-C30 summary score was calculated using all functional and symptom subscales except global QoL and financial impact [23].

### Fatigue

Fatigue was measured with the multidimensional fatigue inventory (MFI-20), a 20-item questionnaire containing five dimensions: general fatigue, physical fatigue, mental fatigue, reduced motivation, and reduced activity [24].

### Anxiety and depression

Anxiety and depression was measured with the validated Dutch version of the hospital anxiety and depression scale (HADS), containing seven items for the depression subscale and seven items for the anxiety subscale [25].

### Physical activity

Physical activity during an average week in the past months was assessed using the Short QUESTIONNAIRE to ASsess Health enhancing physical activity (SQUASH) [26]. A total score was calculated by summing up the minutes per week spent in commuting activities (cycling), leisure time walking and cycling, and sports activities, all with  $\geq 4$  MET.

### UMBRELLA Fit measurements

Patients participating in the exercise intervention visited the UMCU pre- and 12-week post-intervention for additional measurements, i.e., questionnaires, cardiopulmonary exercise testing, intervention acceptance and compliance (see details in the supplement) [18].

### Sample size calculation

A required sample size of 166 patients was estimated based on an expected acceptance rate of 70% in the intervention group and a clinically relevant 10-point difference in global QoL (power of 80% and two-sided alpha of 0.05) [18, 27, 28]. After recruitment of 152 patients, the actual acceptance rate was lower than expected (i.e., 55% instead of 70%) and the sample size was updated, as recommended by Candlish et al. to 260 patients [29]. This sample size adaptation was solely based on the acceptance rate. No interim analysis of the trial outcome was performed.

### Statistical analysis

The statistical analyses were specified in our study protocol, which was approved by the Ethics Committee before recruitment started (NL52062.041.15). The statistical analyses plan remained unchanged, but since the TwiCs design is relatively new and little was known about the analyses methods, methods were further refined throughout the analyses process, which was also one of the aims of the current study and which was stated in the statistical analysis plan.

Baseline characteristics and within-group changes are described for all patients and separately for patients who accepted (including patients who withdrew from the intervention in a later phase) and patients who refused the intervention.

### Within-group changes

Mean changes and corresponding 95%CI in QoL, fatigue, anxiety and depression, and physical activity level from baseline to 6 months follow-up (cohort measurements) and from pre- to 12-week post-intervention (exercise group) were calculated.

### Between-group differences

ITT linear regression analysis was used to assess between-group differences. ITT analysis might lead to an underestimation of the intervention effect because of intervention refusal [30]. Therefore, we also performed instrumental variable IV analysis using the two-stage least squares method to account for possible non-acceptance in the intervention group as a sensitivity analysis [29, 31]. In the first stage, the relation between treatment assignment and treatment acceptance (compliance) was estimated [32]. In the second stage, the effect of the exercise program on the outcome was estimated, using the predicted values from the first stage as an independent variable in a linear regression model. In the ITT and IV analyses, missing values on covariates and baseline measures of the outcome were multiply imputed (15 imputed datasets

using the R mice algorithm [33, 34]), whereas patients with missing outcome values were omitted [35].

In the IV analysis, we could not rule out that intervention refusers were influenced by the offer of the intervention [36]. Therefore, we performed propensity score (PS) analysis as a second sensitivity analysis. Here we estimated the effect in intervention accepters by comparing them to control patients who would have accepted the exercise intervention if offered. To this end, first, propensity scores were estimated for patients in the intervention group using a logistic regression model, i.e., the probability of accepting the intervention, given their observed characteristics (baseline measures of the outcome, age, time since diagnosis, BMI, education). Second, based on the propensity scores, intervention accepters were matched to potential accepters in the control group (i.e., patients who would have accepted the intervention if offered) in a 1:1 ratio without replacement using nearest neighbor matching. In the matched sample, balance of covariates between intervention groups was assessed by means of standardized mean differences and by checking the C-statistic of a refitted propensity score model, indicating the ability of the model to predict treatment status. Finally, linear regression analysis was performed. For PS analysis, missing values of the outcomes were also imputed [33].

All models were adjusted for baseline measures of the outcome, age, time since diagnosis, BMI, and education.

We did not adjust for multiple testing, partly because of multicollinearity between outcomes [37]. Therefore, we reported for all secondary endpoints the effect estimates with 95% confidence intervals and the inferences drawn may not be reproducible.

### Sensitivity analysis

We repeated the ITT and IV analyses where missing values of the outcome also were imputed. Additionally, ITT analysis was repeated replacing the cohort measurements by the pre- and 12-week post-intervention measurements for patients who did not yet start or did not yet complete the intervention before the cohort follow-up measurement (Fig. 3). Furthermore, PS analysis was performed in refusers (i.e., those who were offered the intervention, yet refused it) to check whether there was an effect of offering the intervention in refusers.

We used SPSS version 25.0 or R Statistical Software version 3.5.1. For all models, model fit was checked and was satisfying.

## Results

### Patients and participation

In total, 260 patients were randomly allocated to the intervention ( $n = 130$ ) or control group ( $n = 130$ ; Table 1, Fig. 2).

Of the patients allocated to the intervention arm, 52% accepted the intervention ( $n = 68$ ). Patients who accepted were slightly younger, had a lower BMI, and were higher educated than refusers (Table 1).

### Adherence

Eight patients withdrew from the intervention after a median number of four training sessions (range 1–15). The 60 patients who completed the exercise intervention attended on average 92% (SD = 9.9) of 24 sessions. At baseline, intervention accepters had a mean  $VO_{2peak}$  of 23.8 ml/min/kg (SD = 5.4), which increased, on average, with 1.8 ml/min/kg (95% CI = - 0.4; 0.4) following the exercise intervention.

Timing of the cohort measurements was fixed (Fig. 3). Consequently, about half ( $n = 35/68$ ) of the patients completed the follow-up cohort questionnaire before intervention completion, whereof three patients started the intervention after the follow-up cohort questionnaire. Reasons for a delayed intervention start were planned vacations, physical conditions (e.g., elective heart surgery), family issues, and time constraints. Eleven intervention (16%) accepters did not complete the follow-up questionnaire.

### Retention

The 6-month follow-up cohort questionnaire was returned by 87% of the patients. Patients who did not return this questionnaire were lower educated and had a lower baseline global QoL. The response rate was lowest in intervention refusers (77%) and highest in accepters (93%). Eighty percent of control patients returned the follow-up cohort questionnaire.

### Physical activity

Mean increase in physical activity level from baseline to 6-month follow-up in the intervention group was 61 min/week (95% CI = 26; 97, Table 2). On average, intervention accepters reported an increase of 67 (95% CI = 26; 109). The increase in refusers was comparable (53 min/week, 95% CI = - 11; 117), but not different from zero. Physical activity in the control group did not change substantially (17, 95% CI = - 16; 51).

### Quality of life

At baseline, global QoL was comparable to the Dutch general female population (mean = 76.9, SD = 17.9). Within-group changes are shown in Table 3. Compared to control, the exercise intervention did not result in improved global QoL (ITT: 0.8, 95% CI = - 2.2; 3.8, ES = 0.05; IV: 1.3, 95% CI = - 3.9; 6.5) and the summary score (ITT: 0.8,

**Table 1** Baseline characteristics by randomization group and of patients who either accepted or refused the offer of the intervention

	Control ( <i>n</i> = 130)	Intervention ( <i>n</i> = 130)	Intervention accepted ( <i>n</i> = 68)	Intervention refused ( <i>n</i> = 62)
Age (years), mean (SD)	58.3 (9.5)	58.0 (9.8)	56.6 (9.8)	59.4 (9.6)
Time since cohort inclusion (weeks), mean (SD)	65.9 (13.0)	65.6 (13.1)	66.6 (13.5)	64.5 (12.7)
Body weight (kg), mean (SD)	74.8 (15.2)	74.9 (15.7)	73.6 (13.7)	76.5 (17.8)
BMI, mean (SD)	26.3 (5.0)	26.6 (5.4)	26.0 (4.6)	27.3 (6.2)
Education, <i>n</i> (%)				
Primary or (post-)secondary school	71 (56)	80 (64)	39 (59)	41 (71)
College or university	57 (44)	44 (36)	27 (41)	17 (29)
Smoking, <i>n</i> (%)				
Never	59 (46)	47 (38)	27 (41)	20 (35)
Ex-smoker	47 (37)	59 (48)	31 (47)	28 (48)
Current smoker	22 (17)	18 (15)	8 (12)	10 (17)
Breast cancer stage, <i>n</i> (%)				
In situ/DCIS	16 (14)	20 (17)	12 (19)	8 (14)
Stage I	62 (54)	52 (44)	28 (45)	24 (42)
Stage II	33 (29)	41 (35)	18 (29)	23 (40)
Stage III	4 (4)	6 (5)	4 (7)	2 (4)
Treatment, <i>n</i> (%)				
Chemotherapy	42 (37)	50 (42)	27 (44)	23 (40)
Hormonal therapy	46 (40)	48 (40)	23 (37)	25 (44)

*BMI* body mass index, *DCIS* ductal carcinoma in situ, *SD* standard deviation

No differences between patients who accepted and refused the offer of the intervention were observed

Information on body weight, BMI, education, and smoking available for  $N_{\text{control}} = 128$ ,  $N_{\text{intervention}} = 124$ ,  $N_{\text{accepters}} = 66$ ,  $N_{\text{refusers}} = 58$

Information on breast cancer stage and treatment available for  $N_{\text{control}} = 115$ ,  $N_{\text{intervention}} = 119$ ,  $N_{\text{accepters}} = 62$ ,  $N_{\text{refusers}} = 57$

95% CI = − 1.4; 3.0, ES = 0.06; IV: 1.3, 95% CI = − 2.6; 5.1, Table 4).

In PS analyses, all covariates were well balanced (i.e., standardized mean differences < 0.1). No differences in the QoL measures were found between actual intervention accepters in the intervention arm and potential accepters in the control arm.

## Fatigue

The intervention group improved in general fatigue (MD = − 0.9, 95% CI = − 1.5; − 0.3), physical fatigue (MD = − 1.2, 95% CI = − 1.7; − 0.6), and “reduced activity” (MD = − 0.8, 95% CI = − 1.3; − 0.2) from baseline to 6 months (Table 3). Intervention accepters showed improvements in these three dimensions, while refusers only improved in physical fatigue (Table 3). In the control group, fatigue did not change over time. When using the pre- and 12-week post-intervention outcomes, intervention patients improved on all fatigue dimensions (Supplemental Table 2).

Compared to controls, patients in the intervention group reported larger, but still small, reductions in physical fatigue (ITT: − 1.1, 95% CI = − 1.8; 0.3, ES = − 0.24; IV: − 1.8,

95% CI = − 3.2; − 0.5; PS: − 1.2; 95% CI = − 2.3; − 0.2, Table 4). No between-group differences were found for the other fatigue dimensions.

## Anxiety and depression

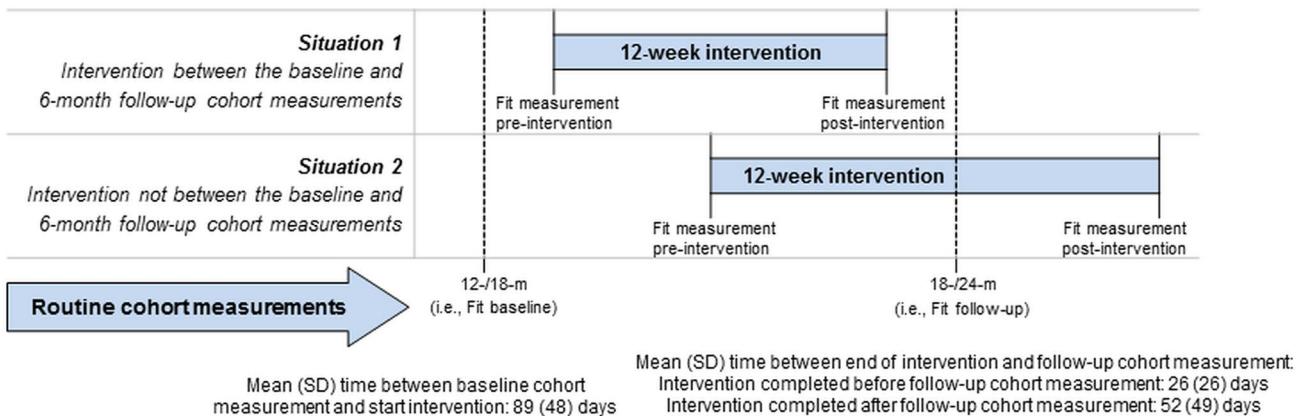
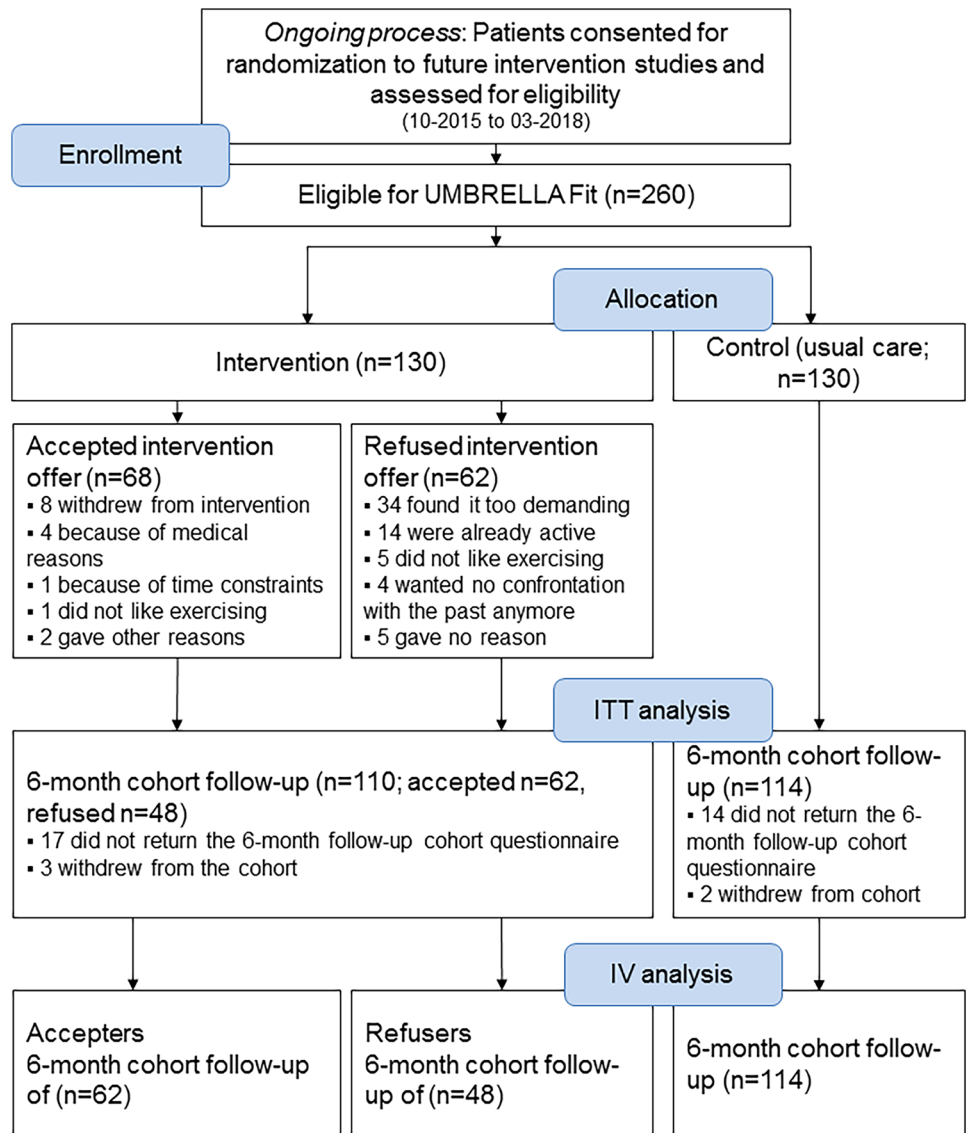
No differences between the intervention and control group were found on anxiety and depression (ITT: − 0.2, 95% CI = − 1.2, ES = − 0.03; 0.8 and 0.0, 95% CI = − 1.0; 0.9, ES = 0.03, respectively).

## Sensitivity analysis

Fifteen percent of the patients in the intervention group ( $n = 20$ ; whereof 6 intervention accepters and 14 intervention refusers) and twelve percent of the patients in the control group had a missing value on the primary outcome ( $n = 16$ ). Repeating the analyses with missing values on the primary outcome imputed as well as replacing the cohort measurements by the pre- and 12-week post-intervention measurements for patients who did not yet start or not yet complete the intervention before the cohort follow-up measurement, yielded comparable results (Supplemental Material).



**Fig. 2** Flow diagram of recruitment, randomization, and follow-up in the UMBRELLA Fit study



**Fig. 3** Timing of the intervention (and hence the pre- and 12-week post measurements) relative to the cohort measurements

**Table 2** Physical activity level at baseline and changes from baseline to 6 months follow-up

Physical activity level, minutes per week	Control		Intervention		Intervention accepted		Intervention refused		Pre- and 12-week post measurements <sup>a</sup>
	<i>n</i>		<i>n</i>		<i>n</i>		<i>n</i>		
Baseline, median (IQR)	127	60 (0; 180)	127	60 (0; 120)	67	60 (0; 140)	60	50 (0; 120)	75 (0; 210)
Change from baseline to 6 months follow-up, mean (95% CI)	110	17 (-16; 51)	108	61 (26; 97)	62	67 (26; 109)	46	53 (-11; 117)	60 (19; 100)

CI confidence interval, IQR interquartile range

<sup>a</sup>UMBRELLA Fit pre- and post-intervention measurements in the patients who accepted the intervention

However, the intervention group reported larger, but still small reductions in general fatigue and reduced activity compared to the control group when using the 12-week post-intervention outcomes ( $-0.9$ , 95% CI =  $-1.7$ ;  $-0.0$  and  $-0.9$ , 95% CI =  $-1.7$ ;  $-0.1$  respectively).

No differences in QoL and fatigue were found between intervention refusers and potential refusers (control group; data not shown).

## Discussion

This study showed no effect on QoL, and a small but beneficial effect of a 12-week exercise intervention on physical fatigue in patients after breast cancer treatment. More than half of the patients performing no or little physical activity started exercising. The TwiCs design provided insights in the effect of refusing an intervention offer. Interestingly, in refusers, we observed slightly higher levels of physical activity, QoL and fatigue compared to the control group, which was not statistically significant when formally tested. This may imply that offering an intervention and actively refusing the offer might induce lifestyle changes.

In contrast to a systematic review reporting positive effects of exercise on QoL in breast cancer patients, we found no effect [8]. In UMBRELLA Fit, eligibility of patients was based on a physically inactive lifestyle and not on a low QoL. QoL of the study population was comparable to the Dutch general female population [38]. Therefore, room for improvement in QoL was limited. In a previous trial from our group in which patients experienced at least three problems on QoL domains, exercise indeed led to a relevant increase in QoL compared to control [40].

In line with a meta-analysis that showed that physical fatigue is the most sensitive fatigue dimension in exercise-oncology trials, we found a positive effect of exercise on physical fatigue [41]. In UMBRELLA cohort patients, QoL decreased during treatment and recovered in the year after treatment, whereas fatigue remained high as compared to a female Dutch reference population, even 18 months after treatment (unpublished observation), and is thus a symptom that indeed needs attention.

## Implications of the TwiCs design for effect estimation

Because patients may refuse the intervention, it is recommended from simulation studies to perform both ITT and IV analyses to take non-acceptance into account when estimating the ‘real’ intervention effect [29, 30]. The ITT analysis showed the effect of offering an exercise intervention, which resembles clinical practice, but it may dilute the treatment effect dependent on the extent of non-acceptance. The IV

**Table 3** Quality of life, fatigue, and anxiety and depression at baseline and within-group changes

	Baseline		Within-group	
	<i>n</i>	Mean (SD)	<i>n</i>	Mean change (95% CI)
EORTC-QLQ-C30—Functional scales <sup>a</sup>				
QLQ-C30 summary score; 89.3 (11.9) <sup>b</sup>				
Control	127	87.2 (11.4)	111	6.4 (4.8; 8.0)
Intervention	125	85.5 (11.5)	107	7.1 (5.6; 8.7)
Intervention accepted	66	85.7 (10.2)	61	7.3 (5.1; 9.5)
Intervention refused	59	85.3 (12.9)	46	6.9 (4.7; 9.2)
UMBRELLA Fit measurement <sup>c</sup>	60	85.1 (9.0)	60	4.1 (2.3; 5.9)
Global quality of life; 76.9 (17.9) <sup>b</sup>				
Control	130	79.0 (17.0)	114	0.4 (− 1.9; 2.8)
Intervention	127	76.6 (17.1)	110	1.9 (− 0.4; 4.2)
Intervention accepted	67	78.0 (17.0)	62	1.2 (− 2.0; 4.4)
Intervention refused	60	75.1 (17.2)	48	2.8 (− 0.6; 6.2)
UMBRELLA Fit measurement <sup>c</sup>	60	76.1 (17.5)	60	8.1 (4.1; 12.0)
Physical functioning; 88.5 (16.0) <sup>b</sup>				
Control	130	87.3 (13.9)	114	0.0 (− 1.7; 1.7)
Intervention	127	86.7 (13.9)	110	− 0.2 (− 2.0; 1.6)
Intervention accepted	67	88.2 (12.4)	62	− 0.1 (− 2.5; 2.3)
Intervention refused	60	85.1 (15.3)	48	− 0.4 (− 3.0; 2.2)
UMBRELLA Fit measurement <sup>c</sup>	60	86.6 (12.7)	60	3.9 (1.8; 6.0)
Role functioning; 86.3 (22.3) <sup>b</sup>				
Control	130	84.7 (20.2)	114	0.1 (− 3.4; 3.7)
Intervention	127	80.6 (23.3)	110	2.3 (− 1.8; 6.3)
Intervention accepted	67	80.1 (23.8)	62	1.1 (− 4.8; 7.0)
Intervention refused	60	81.1 (22.9)	48	3.8 (− 1.4; 9.1)
UMBRELLA Fit measurement <sup>c</sup>	60	84.2 (18.0)	60	5.3 (0.7; 9.9)
Emotional functioning; 86.6 (17.7) <sup>b</sup>				
Control	130	83.5 (19.7)	114	0.0 (− 3.2; 3.2)
Intervention	127	83.7 (17.7)	110	1.8 (− 1.5; 5.0)
Intervention accepted	67	83.3 (17.1)	62	2.0 (− 2.1; 6.1)
Intervention refused	60	84.1 (18.5)	48	1.5 (− 3.7; 6.8)
UMBRELLA Fit measurement <sup>c</sup>	60	81.3 (14.1)	60	7.1 (3.2; 10.9)
Cognitive functioning; 91.6 (15.8) <sup>b</sup>				
Control	130	81.2 (21.5)	114	0.0 (− 2.8; 2.8)
Intervention	127	83.6 (17.9)	110	1.7 (− 1.3; 4.7)
Intervention accepted	67	83.3 (16.9)	62	1.1 (− 3.2; 5.4)
Intervention refused	60	83.9 (19.2)	48	2.4 (− 1.7; 6.6)
UMBRELLA Fit measurement <sup>c</sup>	60	80.0 (20.5)	60	4.7 (0.8; 8.7)
Social functioning; 92.4 (17.3) <sup>b</sup>				
Control	130	88.6 (17.6)	114	2.2 (− 0.6; 5.0)
Intervention	127	89.1 (19.3)	110	0.5 (− 2.5; 3.5)
Intervention accepted	67	91.5 (15.5)	62	0.5 (− 3.2; 4.3)
Intervention refused	60	86.4 (22.7)	48	0.3 (− 4.6; 5.3)
UMBRELLA Fit measurement <sup>c</sup>	60	88.1 (21.7)	60	6.7 (1.1; 12.3)
MFI-20 <sup>d</sup>				
General fatigue				
Control	130	10.6 (4.3)	113	− 0.3 (− 1.0; 0.4)
Intervention	127	11.5 (4.6)	109	− 0.9 (− 1.5; − 0.3)
Intervention accepted	66	11.6 (4.7)	61	− 1.0 (− 1.8; − 0.1)
Intervention refused	61	11.4 (4.7)	48	− 0.8 (− 1.6; 0.0)



**Table 3** (continued)

	Baseline		Within-group	
	<i>n</i>	Mean (SD)	<i>n</i>	Mean change (95% CI)
UMBRELLA Fit measurement <sup>c</sup>	60	11.1 (3.6)	60	− 2.3 (− 3.3; − 1.3)
Physical fatigue				
Control	130	9.6 (4.1)	113	0.0 (− 0.6; 0.5)
Intervention	127	10.4 (4.8)	108	− 1.2 (− 1.7; − 0.6)
Intervention accepted	66	10.2 (4.9)	61	− 1.2 (− 2.1; − 0.4)
Intervention refused	61	10.7 (4.6)	47	− 1.0 (− 1.8; − 0.3)
UMBRELLA Fit measurement <sup>c</sup>	60	10.5 (3.6)	60	− 2.9 (− 3.7; − 2.0)
Mental fatigue				
Control	130	9.2 (4.3)	113	0.5 (− 0.2; 1.1)
Intervention	127	9.1 (4.0)	108	0.4 (− 0.3; 1.0)
Intervention accepted	66	9.3 (4.1)	61	0.0 (− 0.8; 0.9)
Intervention refused	61	8.9 (3.9)	47	0.8 (− 0.1; 1.7)
UMBRELLA Fit measurement <sup>c</sup>	60	10.1 (3.8)	60	− 1.2 (− 1.9; − 0.5)
Reduced motivation				
Control	130	8.5 (3.9)	114	0.0 (− 0.5; 0.6)
Intervention	127	9.1 (4.2)	109	− 0.2 (− 0.7; 0.3)
Intervention accepted	66	8.7 (4.0)	61	− 0.5 (− 1.0; 0.0)
Intervention refused	61	9.6 (4.3)	48	0.2 (− 0.9; 1.2)
UMBRELLA Fit measurement <sup>c</sup>	60	8.4 (3.2)	60	− 1.1 (− 1.9; − 0.2)
Reduced activity				
Control	130	9.3 (4.3)	113	− 0.3 (− 0.9; 0.3)
Intervention	127	10.3 (4.4)	109	− 0.8 (− 1.3; − 0.2)
Intervention accepted	67	10.1 (4.3)	61	− 1.1 (− 1.8; − 0.3)
Intervention refused	60	10.5 (4.5)	48	− 0.4 (− 1.2; 0.5)
UMBRELLA Fit measurement <sup>c</sup>	60	9.8 (4.0)	60	− 2.2 (− 3.2; − 1.2)
HADS <sup>e</sup>				
Anxiety				
Control	127	5.2 (3.2)	111	0.3 (− 0.2; 0.8)
Intervention	126	4.7 (3.2)	110	0.2 (− 0.4; 0.7)
Intervention accepted	66	4.7 (3.0)	62	0.1 (− 0.5; 0.8)
Intervention refused <sup>d</sup>	60	4.7 (3.5)	48	0.2 (− 0.7; 1.1)
Depression				
Control	127	3.1 (3.4)	111	0.2 (− 0.2; 0.6)
Intervention	126	2.9 (3.3)	110	0.0 (− 0.4; 0.4)
Intervention accepted	66	2.6 (2.8)	62	− 0.8 (− 0.5; 0.4)
Intervention refused	60	3.4 (3.7)	48	0.1 (− 0.7; 0.9)

Adjusted for corresponding baseline measures of the outcome, age, time since diagnosis, BMI and education

CI confidence interval; SD standard deviation

<sup>a</sup> Scores ranged from 0 to 100 and a higher score indicated better outcomes

<sup>b</sup> Normative data from the Dutch general female population [38]

<sup>c</sup> UMBRELLA Fit pre- and 12-week post-intervention measurements in the patients who accepted the intervention

<sup>d</sup> Scores ranged from 4 to 20 and a higher score indicated more fatigue

<sup>e</sup> Scores ranged from 0 to 21 and a higher score indicated higher levels of anxious or depressive state

analysis took the chance of non-acceptance into account. To check its robustness, we also performed PS analyses and compared intervention accepters with potential accepters

in the control group. All (sensitivity) analyses yielded comparable results, indicating that the impact of intervention refusal on the effect size seems small. Although no

**Table 4** Effect of the exercise intervention on quality of life, fatigue, and anxiety and depression

	ITT Mean difference (95% CI)	ES <sup>a</sup>	IV Mean difference (95% CI)	PSM Mean difference (95% CI)
EORTC QLQ-C30—Functional scales <sup>b</sup>				
QLQ-C30 summary score				
Control	Ref			
Intervention	0.8 (− 1.4; 3.0)	0.06	1.3 (− 2.6; 5.1)	0.7 (− 2.6; 3.9)
Global quality of life				
Control	Ref			
Intervention	0.8 (− 2.2; 3.8)	0.05	1.3 (− 3.9; 6.5)	0.2 (− 4.2; 4.6)
Physical functioning				
Control	Ref			
Intervention	− 0.4 (− 2.8; 2.1)	− 0.02	− 0.7 (− 5.0; 3.6)	− 0.3 (− 3.6; 2.9)
Role functioning				
Control	Ref			
Intervention	1.2 (− 3.8; 6.1)	0.05	1.9 (− 6.9; 10.8)	− 0.7 (− 8.0; 6.5)
Emotional functioning				
Control	Ref			
Intervention	2.5 (− 1.3; 6.3)	0.15	4.3 (− 2.4; 11.1)	4.1 (− 1.0; 7.2)
Cognitive functioning				
Control	Ref			
Intervention	3.0 (− 0.9; 6.9)	0.15	5.1 (− 1.8; 12.0)	1.0 (− 4.5; 7.2)
Social functioning				
Control	Ref			
Intervention	− 0.0 (− 4.5; 2.7)	− 0.05	− 1.6 (− 8.1; 4.8)	− 0.5 (− 5.2; 4.3)
MFI-20 <sup>c</sup>				
General fatigue				
Control	Ref			
Intervention	− 0.4 (− 1.3; 0.5)	− 0.09	− 0.7 (− 2.2; 0.8)	− 0.2 (− 1.4; 1.0)
Physical fatigue				
Control	Ref			
Intervention	− 1.1 (− 1.8; − 0.3)	− 0.24	− 1.8 (− 3.2; 0.5)	− 1.2 (− 2.3; − 0.2)
Mental fatigue				
Control	Ref			
Intervention	− 0.2 (− 1.1; 0.7)	− 0.04	− 0.3 (− 1.8; 1.2)	− 0.4 (− 1.7; 0.9)
Reduced motivation				
Control	Ref			
Intervention	− 0.2 (− 0.9; 0.5)	− 0.05	− 0.4 (− 1.7; 1.0)	− 0.5 (− 1.5; 0.5)
Reduced activity				
Control	Ref			
Intervention	− 0.3 (− 1.1; 0.5)	− 0.07	− 0.6 (− 1.9; 0.8)	− 0.7 (− 1.8; 0.4)
HADS <sup>d</sup>				
Anxiety				
Control	Ref			
Intervention	− 0.1 (− 0.8; 0.6)	− 0.03	− 0.1 (− 1.3; 1.1)	− 0.2 (− 1.2; 0.8)
Depression				
Control	Ref			
Intervention	0.1 (− 1.0; 1.1)	0.03	0.2 (− 0.8; 1.2)	0.0 (− 1.0; 0.9)

Adjusted for corresponding baseline measures of the outcome, age, time since diagnosis, BMI and education

CI confidence interval; ES standardized effect size; ITT intention-to-treat; IV instrumental variable; PSM propensity score matching

$N_{\text{control}} = 114$ ,  $N_{\text{intervention}} = 110$ ,  $N_{\text{PSMaccepters/arm}} = 67$ ,  $N_{\text{PSMrefusers/arm}} = 62$  (some patients do not have an outcome score on a particular scale or domain)

<sup>a</sup>According to Cohen [39],  $ES \leq 0.2$  = small effect,  $ES 0.2-0.5$  = medium effect,  $ES \geq 0.8$  = large effect

<sup>b</sup>Scores ranged from 0 to 100 and a higher score indicated better outcomes

<sup>c</sup>Scores ranged from 4 to 20 and a higher score indicated more fatigue

<sup>d</sup>Scores ranged from 0 to 21 and a higher score indicated higher levels of anxious or depressive state

differences in outcomes were found between patients who refused the intervention and potential refusers in the control group, we cannot exclude that there was an effect of offering the intervention on outcomes, which then also affects the IV and PS analyses. However, analyses were not powered to detect an effect in refusers; therefore, these results are explorative.

Due to the design, we experienced that it was sometimes challenging to schedule the intervention in between the two cohort measurements. Consequently, 35 intervention accepters (of 68, 51%) had not yet completed the intervention or started ( $n = 3/68$ , 4%) with the intervention when the endpoint was evaluated. Therefore, the effects may be underestimated. As a sensitivity analysis, we replaced the cohort measurements by the pre- and 12-week post-intervention measurements for these patients and this yielded slightly larger effects. We also considered using the first cohort measurement after the study end for patients who did not yet complete the intervention before the planned cohort outcome measurement. These could then be matched to random control patients. However, this could introduce a risk of immortal time bias since both groups need to survive until the extended measurement point. Intervention and control patients could be matched on baseline variables, yet this would require selection based on post-intervention variables (notably survival until end of follow-up), thus potentially introducing a selection bias. Given these considerations, we did not conduct these sensitivity analyses. We recommend being stringent in intervention planning and taking a suitable follow-up period when designing a TwiCs.

We observed that the 6-month changes were smaller than the 12-week post-intervention changes. First, as described above, not every patient completed the intervention before the endpoint was evaluated. Second, cohort measurements were not obtained in a trial setting, whereas patients may have completed the 12-week post-intervention questionnaire with their trial participation in mind. Third, the 6-month cohort outcomes might present a ‘long-term’ effect of the intervention. Therefore, long-term cohort outcomes might be representative for the (long-term) real world effects.

### Implications for clinical practice

Because the TwiCs consent procedure better reflects clinical practice than in conventional trials, this study provided important insight into intervention uptake and reasons for refusal. Because we used the TwiCs design, we learned that more than half of the patients performing no or little physical activity started exercising and subsequent adherence was high. Therefore, health care providers should not hesitate to motivate patients to become physically active. Yet, almost half of the patients refused the offer of the exercise intervention, even when it was free of costs. Based on their reasons

for refusal, one can consider adaptations to, or alternatives for the intervention, to reinforce uptake in intervention refusers, e.g., offering these patients a home-based exercise program.

### Conclusion

In this TwiCs, more than half of the inactive patients with breast cancer accepted the offer of a 12-week supervised exercise intervention. We found no effects of offering an exercise program on QoL, which was already high at baseline. For physical fatigue, we found small but beneficial exercise effects. Applying the TwiCs design appeared feasible. The use of the TwiCs design could dilute effect sizes because of intervention refusal. In this TwiCs, the impact of intervention refusal seemed small because of consistent results of the ITT and IV analysis (taking non-compliance into account). In addition, the use of cohort measurements for effect estimation may have diluted the effect size. Therefore, we recommend careful intervention planning in-between the cohort measurements that will be used as pre- and post-intervention measurement.

**Acknowledgements** The authors thank the study team of the UMBRELLA cohort at the department of Radiotherapy, Division of Imaging and Oncology, and Julius Center for Health Sciences and Primary Care of the University Medical Center Utrecht (UMCU) for running the cohort and collecting the data. The authors also thank the Department of Rehabilitation, Physical Therapy Science and Sports, University Medical Center Utrecht (UMCU) for the collaboration and their assistance.

**Author contributions** AMM was responsible for the acquisition of funding for UMBRELLA Fit, and HMV and CHG for the UMBRELLA cohort. AMM and EMM designed the study and were responsible for the general supervision of the study. RHHG, CHG, DHJB, PHMP, and HMV were involved in designing the study. RG was responsible for the acquisition of data. RG, RHHG, SGE, AMM, and EMM were involved in the analysis. RG drafted the manuscript and all authors critically revised the paper. All authors read, provided feedback, and approved the final manuscript.

**Funding** This work was supported by a Veni grant from the The Netherlands Organisation for Health Research and Development (ZonMw 016.156.050). In addition, funding was provided by the department of Radiotherapy, Imaging Division of the University Medical Center Utrecht (UMCU), and the department of Clinical Epidemiology, Julius Center for Health Sciences and Primary Care, UMCU.

**Data availability** The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Code availability** The R code is available from the corresponding author upon reasonable request.

## Declarations

**Conflict of interest** The authors declare no potential conflicts of interest.

**Ethics approval** This study was conducted according to the principles of the Declaration of Helsinki (Forteleza, October 2013: <http://www.wma.net/en/30publications/10policies/b3/>) and in accordance with the Medical Research Involving Human Subjects Act (WMO). The study has been approved by the Medical Ethics Committee of the University Medical Centre Utrecht (15-166/D).

**Consent to participate** Within the TwiCs design, a stage-informed consent procedure is applied [20]. In the first stage, at cohort entry, patients are asked for consent to participate in the cohort study (collection of medical and patient reported outcomes) and for consent to be randomized for future intervention studies within the cohort. Next, patients who are randomly selected for the intervention group in a RCT within the cohort are asked for consent to participate in the intervention study.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Servaes P, Verhagen C, Bleijenberg G (2002) Fatigue in cancer patients during and after treatment: prevalence, correlates and interventions. *Eur J Cancer* 38(1):27–43. [https://doi.org/10.1016/S0959-8049\(01\)00332-x](https://doi.org/10.1016/S0959-8049(01)00332-x)
- Weis J (2011) Cancer-related fatigue: prevalence, assessment and treatment strategies. *Expert Rev Pharmacoecon Outcomes Res* 11(4):441–446. <https://doi.org/10.1586/erp.11.44>
- Arndt V, Stegmaier C, Ziegler H, Brenner H (2006) A population-based study of the impact of specific symptoms on quality of life in women with breast cancer 1 year after diagnosis. *Cancer* 107(10):2496–2503. <https://doi.org/10.1002/cncr.22274>
- Buffart LM, Kalter J, Sweegers MG, Courneya KS, Newton RU, Aaronson NK, Jacobsen PB, May AM, Galvao DA, Chinapaw MJ et al (2017) Effects and moderators of exercise on quality of life and physical function in patients with cancer: An individual patient data meta-analysis of 34 RCTs. *Cancer Treat Rev* 52:91–104. <https://doi.org/10.1016/j.ctrv.2016.11.010>
- Cramp F, Byron-Daniel J (2012) Exercise for the management of cancer-related fatigue in adults. *Cochrane Database Syst Rev* 11:CD006145. <https://doi.org/10.1002/14651858.CD006145.pub3>
- Fong DY, Ho JW, Hui BP, Lee AM, Macfarlane DJ, Leung SS, Cerin E, Chan WY, Leung IP, Lam SH et al (2012) Physical activity for cancer survivors: meta-analysis of randomised controlled trials. *BMJ* 344:e70. <https://doi.org/10.1136/bmj.e70>
- Kessels E, Husson O, van der Feltz-Cornelis CM (2018) The effect of exercise on cancer-related fatigue in cancer survivors: a systematic review and meta-analysis. *Neuropsychiatr Dis Treat* 14:479–494. <https://doi.org/10.2147/NDT.S150464>
- Lahart IM, Metsios GS, Nevill AM, Carmichael AR (2018) Physical activity for women with breast cancer after adjuvant therapy. *Cochrane Database Syst Rev* 1:011292. <https://doi.org/10.1002/14651858.CD011292.pub2>
- Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, Tunis S, Bergel E, Harvey I, Magid DJ et al (2009) A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 62(5):464–475. <https://doi.org/10.1016/j.jclinepi.2008.12.011>
- Steins Bisschop CN, Courneya KS, Velthuis MJ, Monninkhof EM, Jones LW, Friedenreich C, van der Wall E, Peeters PH, May AM (2015) Control group design, contamination and drop-out in exercise oncology trials: a systematic review. *PLoS ONE* 10(3):e0120996. <https://doi.org/10.1371/journal.pone.0120996>
- Lindstrom D, Sundberg-Petersson I, Adami J, Tonnesen H (2010) Disappointment and drop-out rate after being allocated to control group in a smoking cessation trial. *Contemp Clin Trials* 31(1):22–26. <https://doi.org/10.1016/j.cct.2009.09.003>
- McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, Elbourne DR, Francis D, Garcia J, Roberts I et al (2006) What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials* 7:9. <https://doi.org/10.1186/1745-6215-7-9>
- Relton C, Torgerson D, O’Cathain A, Nicholl J (2010) Rethinking pragmatic randomised controlled trials: introducing the cohort multiple randomised controlled trial design. *BMJ* 340:c1066. <https://doi.org/10.1136/bmj.c1066>
- Gal R, Monninkhof EM, van Gils CH, Groenwold RHH, van den Bongard D, Peeters PHM, Verkooijen HM, May AM (2019) The trials within cohorts design faced methodological advantages and disadvantages in the exercise oncology setting. *J Clin Epidemiol* 113:137–146. <https://doi.org/10.1016/j.jclinepi.2019.05.017>
- Couwenberg AM, Burbach JPM, May AM, Berbee M, Intven MPW, Verkooijen HM (2020) The trials within cohorts design facilitated efficient patient enrollment and generalizability in oncology setting. *J Clin Epidemiol* 120:33–39. <https://doi.org/10.1016/j.jclinepi.2019.12.015>
- Couwenberg AM, Burbach JPM, Berbee M, Lacle MM, Arensman R, Raicu MG, Wessels FJ, Verdult J, Roodhart J, Reerink O et al (2020) Efficacy of dose-escalated chemoradiation on complete tumor response in patients with locally advanced rectal cancer (RECTAL-BOOST): a phase 2 randomized controlled trial. *Int J Radiat Oncol Biol Phys* 108(4):1008–1018. <https://doi.org/10.1016/j.ijrobp.2020.06.013>
- Pielkenrood BJ, van der Velden JM, van der Linden YM, Bartels MMT, Kasperts N, Verhoeff JJC, Eppinga WSC, Gal R, Verlaan JJ, Verkooijen HML (2020) Pain response after stereotactic body radiation therapy versus conventional radiation therapy in patients with bone metastases—a phase 2 randomized controlled trial within a prospective cohort. *Int J Radiat Oncol Biol Phys*. <https://doi.org/10.1016/j.ijrobp.2020.11.060>
- Gal R, Monninkhof EM, Groenwold RHH, van Gils CH, van den Bongard D, Peeters PHM, Verkooijen HM, May AM (2017) The effects of exercise on the quality of life of patients with breast cancer (the UMBRELLA Fit study): study protocol for a randomized controlled trial. *Trials* 18(1):504. <https://doi.org/10.1186/s13063-017-2252-5>
- Young-Afat DA, van Gils CH, van den Bongard H, Verkooijen HM (2017) The utrecht cohort for multiple breast cancer intervention studies and long-term evaluation (UMBRELLA): objectives, design, and baseline results. *Breast Cancer Res Treat* 164(2):445–450. <https://doi.org/10.1007/s10549-017-4242-4>
- Young-Afat DA, Verkooijen HA, van Gils CH, van der Velden JM, Burbach JP, Elias SG, van Delden JJ, Relton C, van Vulpen M, van der Graaf R (2016) Brief report: staged-informed consent in the

- cohort multiple randomized controlled trial design. *Epidemiology* 27(3):389–392. <https://doi.org/10.1097/EDE.0000000000000435>
21. <https://www.wcrf.org/dietandcancer/recommendations/be-physically-active>
  22. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JC et al (1993) The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 85(5):365–376. <https://doi.org/10.1093/jnci/85.5.365>
  23. Giesinger JM, Kieffer JM, Fayers PM, Groenvold M, Petersen MA, Scott NW, Sprangers MA, Velikova G, Aaronson NK, Group EQoL (2016) Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust. *J Clin Epidemiol* 69:79–88. <https://doi.org/10.1016/j.jclinepi.2015.08.007>
  24. Smets EM, Garssen B, Cull A, de Haes JC (1996) Application of the multidimensional fatigue inventory (MFI-20) in cancer patients receiving radiotherapy. *Br J Cancer* 73(2):241–245. <https://doi.org/10.1038/bjc.1996.42>
  25. Osborne RH, Elsworth GR, Sprangers MA, Oort FJ, Hopper JL (2004) The value of the hospital anxiety and depression scale (HADS) for comparing women with early onset breast cancer with population-based reference women. *Qual Life Res* 13(1):191–206. <https://doi.org/10.1023/B:QURE.0000015292.56268.e7>
  26. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D (2003) Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. *J Clin Epidemiol* 56(12):1163–1169
  27. Borm GF, Fransen J, Lemmens WA (2007) A simple sample size formula for analysis of covariance in randomized clinical trials. *J Clin Epidemiol* 60(12):1234–1238. <https://doi.org/10.1016/j.jclinepi.2007.02.006>
  28. Osoba D, Rodrigues G, Myles J, Zee B, Pater J (1998) Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 16(1):139–144. <https://doi.org/10.1200/JCO.1998.16.1.139>
  29. Candlish J, Pate A, Sperrin M, van Staa T, GetReal Work P (2017) Evaluation of biases present in the cohort multiple randomised controlled trial design: a simulation study. *BMC Med Res Methodol* 17(1):17. <https://doi.org/10.1186/s12874-017-0295-7>
  30. Pate A, Candlish J, Sperrin M, Van Staa TP, GetReal Work P (2016) Cohort multiple randomised controlled trials (cmRCT) design: efficient but biased? A simulation study to evaluate the feasibility of the Cluster cmRCT design. *BMC Med Res Methodol* 16(1):109. <https://doi.org/10.1186/s12874-016-0208-1>
  31. van der Velden JM, Verkooijen HM, Young-Afat DA, Burbach JP, van Vulpen M, Relton C, van Gils CH, May AM, Groenwold RH (2017) The cohort multiple randomized controlled trial design: a valid and efficient alternative to pragmatic trials? *Int J Epidemiol* 46(1):96–102. <https://doi.org/10.1093/ije/dyw050>
  32. Huang FL (2018) Using instrumental variable estimation to evaluate randomized experiments with imperfect compliance. *Pract Assess Res Eval*. <https://doi.org/10.7275/k0p6-yj16>
  33. van Buuren S (2018) Flexible imputation of missing data. CRC Press Book, USA, p 2018
  34. van Buuren S, Groothuis-Oudshoorn K (2011) Mice: multivariate imputation by chained equations in R. *J Stat Softw* 45(3):1–67
  35. Groenwold RH, Moons KG, Vandenbroucke JP (2014) Randomized trials with missing outcome data: how to analyze and what to report. *CMAJ* 186(15):1153–1157. <https://doi.org/10.1503/cmaj.131353>
  36. Dunn G, Maracy M, Tomenson B (2005) Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. *Stat Methods Med Res* 14(4):369–395. <https://doi.org/10.1191/0962280205sm403oa>
  37. Groenwold RHH, Goeman JJ, Cessie SL, Dekkers OM (2021) Multiple testing: when is many too much? *Eur J Endocrinol* 184(3):E11–E14. <https://doi.org/10.1530/EJE-20-1375>
  38. Mols F, Husson O, Oudejans M, Vlooswijk C, Horevoorts N, van de Poll-Franse LV (2018) Reference data of the EORTC QLQ-C30 questionnaire: five consecutive annual assessments of approximately 2000 representative Dutch men and women. *Acta Oncol* 57(10):1381–1391. <https://doi.org/10.1080/0284186X.2018.1481293>
  39. Cohen J (1988) *Statistical power analysis for the behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ
  40. May AM, Korstjens I, van Weert E, van den Borne B, Hoekstra-Weebers JE, van der Schans CP, Mesters I, Passchier J, Grobbee DE, Ros WJ (2009) Long-term effects on cancer survivors quality of life of physical training versus physical training combined with cognitive-behavioral therapy: results from a randomized trial. *Support Care Cancer* 17(6):653–663. <https://doi.org/10.1007/s00520-008-0519-9>
  41. van Vulpen JK, Peeters PH, Velthuis MJ, van der Wall E, May AM (2016) Effects of physical exercise during adjuvant breast cancer treatment on physical and psychosocial dimensions of cancer-related fatigue: a meta-analysis. *Maturitas* 85:104–111. <https://doi.org/10.1016/j.maturitas.2015.12.007>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.