Integrated Short-term Palliative Rehabilitation in Incurable Cancer (INSPIRE) Trial

Integrated Short-term Palliative Rehabilitation to improve quality of life and equitable care access in incurable cancer: A multi-national randomised controlled trial

Statistical Analysis Plan

Version 1.0: 30/10/2024

ISRCTN: 16705336

CONTENTS

ABBREVIATIONS:

eCRF	Electronic Case Report Form
CTU	Clinical Trials Unit
CACE	Complier Average Causal Effects
DMC	Data Monitoring Committee
ECOG	Eastern Cooperative Oncology Group
EMA	European Medicines Agency
HEAP	Health Economic Analysis Plan
ICH	International Conference on Harmonization
INSPIRE	INtegrated Short-term PallIative REhabilitation in incurable cancer
ISRCTN	International Standard Randomised Clinical Trials Number
ITT	Intention to treat
KCL	King's College London
LME	Linear mixed effects model
NHS	National Health Service
RCT	Randomised Control Trial
REC	Research Ethics Committee
SAP	Statistical Analysis Plan
SOP	Standard Operating Procedure
TSC	Trial Steering Committee
UK	United Kingdom
WP	Work Package

INVESTIGATORS:

	Name and E-mail	Address
Trial statistician	Joana Vasconcelos joana.vasconcelos@kcl.ac.uk	Nightingale-Saunders Clinical Trials & Epidemiology Unit Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care; King's College London, Bessemer Road, London SE5 9PJ
Senior statistician	Professor Toby Prevost toby.prevost@kcl.ac.uk	Nightingale-Saunders Clinical Trials & Epidemiology Unit Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care; King's College London, Bessemer Road, London SE5 9PJ
Chief Investigator	Professor Matthew Maddocks matthew.maddocks@kcl.ac.uk	Cicely Saunders Institute of Palliative Care, Policy & Rehabilitation Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care King's College London, Bessemer Road, London SE5 9PJ
Trial Manager	Sylvia Wilczynska sylvia.1.wilczynska@kcl.ac.uk	King’s Clinical Trial Unit; King’s College London PO64, Room M2.06,Institute of Psychiatry, 16 De Crespigny Park, London, SE5 8AF
Chair of TSC	Dr. Nichola Gale galeNS@cardiff.ac.uk	Cardiff University, Cardigan House, University Hospital of Wales, Heath Park, Cardiff, CF14 4WX
Chair of DMC	Professor Sarah Walker rmjlasw@ucl.ac.uk	University College London, MRC Clinical Trials Unit at UCL, 90 High Holborn, London, WC1V 6LJ, United Kingdom
Sponsors	Professor Bashir Al-Hashimi vpri@kcl.ac.uk Dr Guillaume Economos guillaume.economos@chu-lyon.fr	Room 8.11, 8th Floor Melbourne House, 44-46 Aldwych, London, WC2B 4LL, United Kingdom Hospices Civils de Lyon Direction de la Recherche en santé, Direction de l’Innovation, 6 rue Alexander Fleming, Lyon 69007, France

1 Introduction

1.1 Purpose and scope of the statistical analysis plan

The purpose of this Statistical Analysis Plan is to set out the study objectives and hypotheses, and the analytical approaches and procedures necessary to address these for the main trial paper and to provide guidance for further research reported in other papers, promoting consistent approaches and methods.

The SAP in intended primarily for the paper that will report the effectiveness results of the INSPIRE trial, though will provide guidance for further research reported in other papers, by promoting similar approaches and methods where possible, sensible and relevant.

As there can typically be more than one analytical approach to address a hypothesis, there is the potential for different results to be produced from using alternative approaches, alternative methods, alternative outcome definitions and the alternative data that may be involved. These differences can be influential, for example, when results are of borderline statistical significance.

Therefore, this plan records those decisions that can be made about study hypotheses, outcome definitions and statistical procedures, along with their basis and the appropriateness of the assumptions required for their use, in advance of the main trial analysis.
Changes within any subsequent versions of the plan prior to analysis will be dated, with the basis for the changes reasoned, and recorded within the plan.

It is not intended that the strategy set out in the plan should prohibit sensible practices. However, the principles established in the plan will be followed as closely as possible when analysing and reporting the trial.

1.2 Derivation of the statistical analysis plan

The present statistical analysis plan was derived from the trial protocol, by the trial statistician, Joana Vasconcelos, with the supervision of the senior statistician, Professor Toby Prevost. The trial statistician is responsible for developing the SAP as well as for carrying out the statistical analysis for interim data monitoring and final statistical reporting of the trial. The senior statistician will review and revise the SAP and give an overall verification of the analysis throughout the study, in keeping with the Standardised Operating Procedures (SOPs) of the King's Clinical Trials Unit, including the SOP for developing the Statistical Analysis Plan.

The formation of this Plan has drawn on statistical guidance from: the ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trial E9, E9(R1) addendum, and E3 [1], the CONSORT statement for the transparent reporting trials [2] and the Committee for Medicinal Products for Human Use (formerly known as the Committee for Proprietary Medicinal Products) report[3].

The trial statistician will write the first draft version of the plan and after revision by the senior statistician the plan will be filed as draft version number 0.1. The plan will then be discussed with the Chief Investigator and other study members for further input and filed as draft version number 0.2. The plan will then be sent to the DMC and TSC for final approvals and saved as approved version 1.0.

2 Description of the INSPIRE trial

The following text on the trial background was taken from the study protocol v1.0, therefore references are omitted from this section. This palliative rehabilitiation intervention trial was deemed to be importantly needed as it was competitively funded by EU Horizon.

2.1 Background to the condition (cancer) and importance of palliative rehabilitation

Cancer is one of the main causes of illness, burden and death in Europe, and also a major and growing contributor to disability (loss of function). Recent global estimates suggest a loss of 382 disability-adjusted life years per 1000 individuals. Disability is a poorly recognized and undertreated consequence of incurable cancer. Over time, loss of function results in people not being able to continue with valued roles and routines, to manage usual household and social activities, and to self-care. One-third of adults with cancer require assistance to perform basic activities like washing and dressing, and half need help with extended activities like shopping and transportation. Disability reduces quality of life and well-being. Disability related to daily activity is closely related to unplanned hospital admissions and mortality.

Palliative rehabilitation empowers people with incurable conditions to actively manage their condition themselves, enabling them to live fully and enjoy the best health-related quality of life possible, including cancer towards the end of life. It aims to reduce symptoms and help people to stay independent and socially active. WHO policy on Universal Health Coverage states both rehabilitation and palliative care as essential, quality health services. While integrated rehabilitation has been achieved for people with chronic respiratory, cardiac and stroke conditions, this is not the case for people with cancer, especially those living with incurable disease. Access to palliative care services has increased but access to rehabilitation remains varied.

This trial evaluates a rehabilitation intervention that has been designed to meet the needs of people living with advanced cancer. The study is taking place in countries across Europe, and we plan to recruit 340 patients from hospitals. We aim to find out if and how the rehabilitation intervention affects the people who take part in the study. We will also study how it fits in with current healthcare services.

2.2 Description of the intervention

The intervention being tested is Integrated Short-term Palliative Rehabilitation.
It comprises up to 3 manualised sessions (face to face and/or remotely (via telephone or video call) delivered by a rehabilitation practitioner (typically a physiotherapist or occupational therapist). Core components focus on (i) self-management of symptoms, (ii) physical activities and fitness, and (iii) social participation, with explicit use of behaviour change techniques with goal setting and action planning.

The rehabilitation practitioner works in partnership with the person with incurable cancer, and those important to them, to support and optimise their function. Sessions focus on outcomes each person has said are important to them. The rehabilitation practitioner attends to practical, physical, emotional, psychological, and existential concerns impacting on function, either directly within the intervention or indirectly through onward referral. The intervention allows for individual tailoring and flexibility in location, timing and frequency of sessions and content over a 7-week intervention period. Participants can receive a minimum of two rehabilitation sessions and a maximum of three rehabilitation sessions.

It is delivered in addition to any usual services delivered by the participant's oncology team and palliative care team.

2.3 Description of the comparator

Unrestricted usual care, as determined by the healthcare system in the participating countries, within oncology, palliative care, other hospital services or health services in the community and medical practitioner(s) in charge of their care. This will include usual referral to any existing rehabilitation services.

2.4 The target population and the eligibility criteria

The target population, to which inferences from the end of this trial are intended to generalise, is the population of adult patients with incurable solid cancer.

The protocol provides the following eligibility criteria for the study:

Inclusion criteria:

-Aged 18 years or older.
-Diagnosis of incurable solid cancer: lung, colorectal, breast, prostate or other, irrespective of timing in relation to any oncology or palliative care treatments
-Eastern Cooperative Oncology Group performance status 2-3
-Able to provide informed consent and complete trial assessments in available languages.

Exclusion criteria :

-Blood cancers: Leukaemia, Lymphoma, Myelodysplastic Syndromes (MDS), Myeloproliferative Disorder (MPD), Multiple Myeloma.
-Currently receiving specialist rehabilitation for their cancer or co-morbidity-related dysfunction, or received within the two weeks prior to consent.
-Clinician rated prognosis of less than 3 months.

2.5 Principal research objectives

Principal Trial objective

The principal objective of the INSPIRE trial is to assess the clinical effectiveness of palliative rehabilitation over 8 weeks, on health-related quality of life for patients with incurable solid cancer compared to usual care.

Secondary Objectives

i) To assess the effectiveness of palliative rehabilitation over 8 weeks on disability, symptom burden and goal attainment for patients with incurable cancer.

ii) To assess the cost effectiveness of palliative rehabilitation in terms of the changes in the primary outcome measure of quality of life, FACT-G, and to present cost-utility estimates.

iii) To assess cost effectiveness from a health care and societal perspective, focusing on hospital treatment and care costs, ambulatory care costs and cost to informal caregivers over 8 weeks.

iv) To identify which participant characteristics are associated with beneficial randomised intervention effect on quality of life focusing on; sex, gender, age, diagnosis (locally advanced or metastatic disease), performance status, and other subgroup factors.

v) To determine equity, access and patient experience of the intervention, across different cultures, socio- economic and other groups, considering gender, age, religious, cultural and personal beliefs.

vi) To evaluate whether the palliative rehabilitation intervention was successfully implemented and identify factors contributing to successful integration with existing services.

Objectives ii), iii), v) and vi) are not part of this statistical analysis plan for trial effectiveness, and these are covered in the broader coverage of plans for analysis accompanying this plan.

3 Trial design

This is a non-blinded multinational, phase 3, parallel randomised controlled trial. Participants will be randomized in a 1:1 ratio to the arms and will be followed up at weeks 4, 8 and 16.

3.1 Trial arms and blinding

Trial arms

Intervention: The intervention being tested is an integrated short-term palliative rehabilitation. It comprises up to 3 manualised sessions (face to face and/or by telephone) delivered by an expert rehabilitation practitioner (typically a physiotherapist or occupational therapist).

Comparator: Unrestricted usual care, as determined by the healthcare system in the participating countries, within oncology, palliative care, other hospital services or health services in the community and medical practitioner(s) in charge of their care.

Different labels are used within the plan as terminology in describing the control arm, such ‘comparator’ and ‘usual-care’.

Blinding

The trial is Open Label. After allocation to a study arm, both the participant and those delivering intervention or control are aware of the study arm.
The trial statistician will have access to data extracts from the trial databases to provide interim data monitoring reports from accumulating unblinded outcome data that is required for reporting to the DMC. The senior trial statistician will not see each participant’s trial arm, but only aggregated data in veryfiying interim trial reports. Both statisticians have been required to attend open and closed sessions of the DMC meetings.

Given limited blinding, this supports the adopted practice of the development of an early statistical analysis plan, and transparent reporting of any subsequent changes with reasoning.

3.2 Method of allocation to arm

The individuals will be randomised to one of the treatment arms. Randomisation will be done in an equal 1:1, allocation ratio to the two arms.

Randomisation will use the method of minimisation incorporating a random element, balancing the following factors to guard against chance bias in patient allocation for prognostic factors:

i) Trial participant's country: England, Scotland, France, Italy, Denmark, Norway, Czech Republic and any other countries that may join the trial;
ii) Baseline FACT-G score (<=64, 65-79, 80+);
iii) ECOG performance status (2, 3).

3.3 Relative timing of randomisation

Randomisation will be via a bespoke web based randomisation system hosted at the King's CTU on a secure server. Once a participant enters the study, and their data is entered into the eCRF, they will be allocated a unique study PIN. This, along with their date of birth and initials (for those countries where this is available) will be used to identify the participant and their data throughout the study. As randomisation is undertaken by the King's CTU, it is independent of the trial managers, statisticians, investigators and other study staff. The allocated arm for a participant is therefore concealed up until the point of allocation, as recommended in the CONSORT statement.

3.4 Trial dates and duration

If not terminated earlier, the expected duration of the trial is 24 months from opening to recruitment of the first participant to final assessments of all trial participants, cleaning and locking of the trial database.

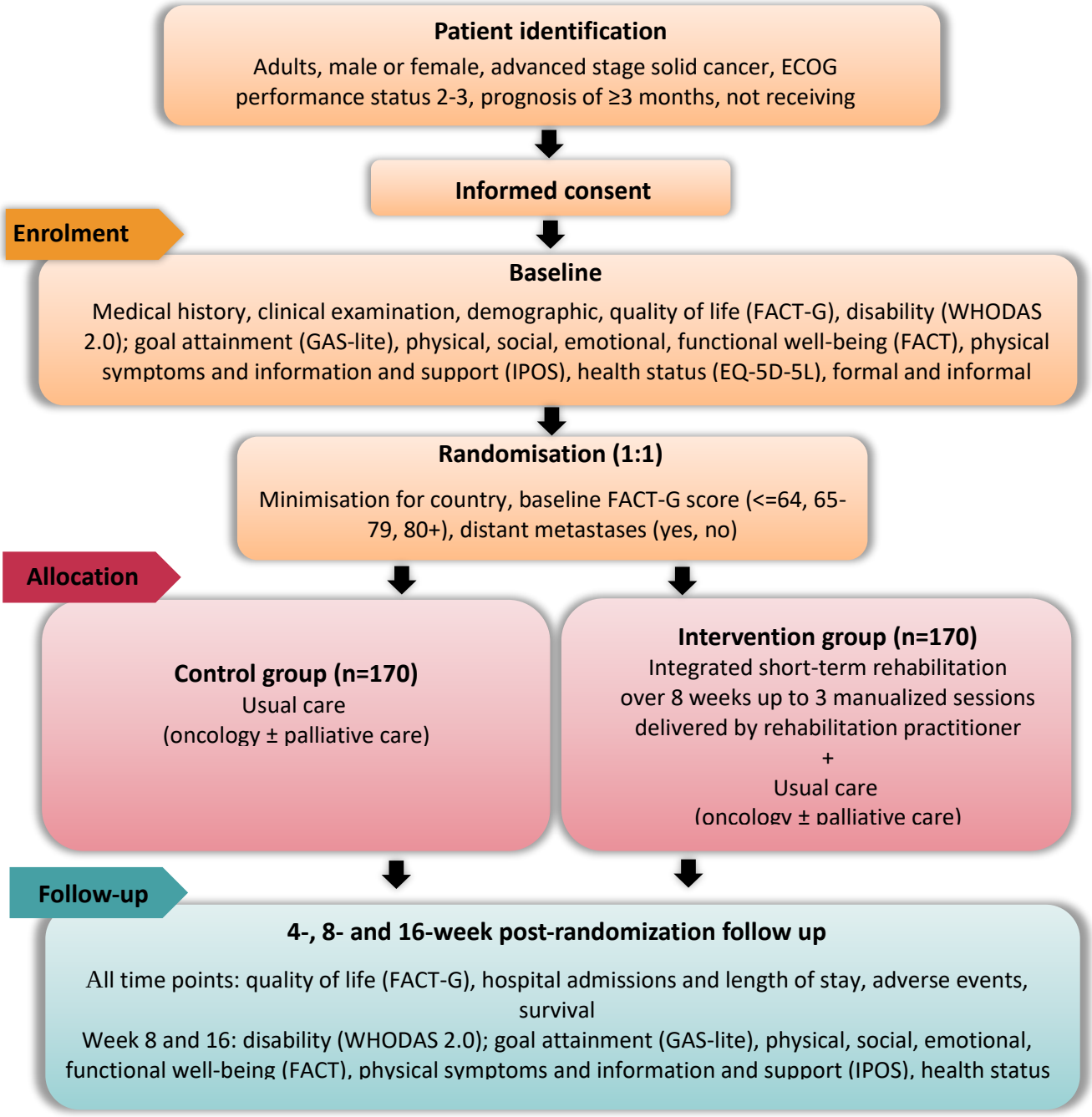
Recruitment starts in May 2024.

3.5 Participant follow-up

Each study subject will participate in the trial from the day that the they give informed consent to their last final visit at 16 weeks and followed up using medical record data until 28 weeks. Participants will be seen at baseline, 4, 8 and 16 weeks. Intervention participants will be also contacted for rehabilitation intervention visits in the first 7 weeks.

3.6 Flow diagram

The flow diagram of the study is shown below. IT will include, by arm, the number randomised, who comprise the intention to treat population, and the number followed-up to be in the analyses of the primary outcome as well as the main reasons for missing data by stages of the trial.



4 Trial measures

4.1 Baseline

As written in protocol section 6.5.3, the socio-demographic information collected at baseline consists of: gender, age, relationship status, living situation, having children, educational level, employment status, financial situation, geographical access to secondary health care, religious status, social support from family or friends, ongoing stressors, perceived discrimination by health care system and others, and health confidence. Ethnicity will also be collected in the UK.

Relevant medical history, comorbidities, clinical diagnosis, date of diagnosis, current treatment, body mass index, weight change, nutrition and physical activity history and blood test results will also be collected as described in protocol section 6.5.4.
Besides this, the scales listed in section 4.3 below, are also collected at baseline.

Of these measurements, the following ones are highlighted to be used in the analysis of the equity, inclusivity and access evaluation work package (WP5):

1. Gender (Woman, Man, Other, Prefer not to answer)
2. Age (<65; ≥65)
3. Diagnosis (locally advanced or metastatic disease)
4. ECOG performance status (2, 3)
5. Country (7 countries currently)
6. Living situation
7. Dependents with care needs (from question – 'Helping at least one sick, limited, or frail family member or friend on a regular basis', with *No* versus *Yes* answers)
8. Health confidence (using either the score or the four items including *I know enough about my illness and treatment plan*; *I can look after my health*; *I can get the right help if I need it*; *I am involved in decisions about me*)

4.2 Primary outcome

The primary outcome is health-related quality of life over the last 7 days as assessed by the Functional Assessment of Cancer Therapy General (FACT-G) scale total score (0=worst quality of life to 108=best quality of life) [4] at 8 weeks after randomisation.

4.3 Secondary outcomes

The secondary outcome measures are listed as follows:

i. Functional assessment:
    FACT-G at 4 and 16 weeks
ii. Disability:
    World Health Organization Disability Assessment Schedule (WHODAS 2.0) at 8 and 16 weeks
iii. Symptoms:
    Integrative Palliative Care Outcome Scale – Physical Symptoms (IPOS-S) at 8 and 16 weeks
iv. Goal attainment:
    Goal attainment scale (GAS-Light) at 8 and 16 weeks
v. Economic reported outcomes (this is detailed in the health economics analysis plan)
    Client Service Receipt Inventory at 8 and 16 weeks

The analysis plan for the latter as well as the implementation outcomes (see INSPIRE protocol) are not covered in this plan.

4.4 Adverse events

Adverse events will be recorded in a log, stating the start/end date, severity, serious adverse event, relatedness to the rehabilitation intervention, impact on the ability to receive rehabilitation and outcome (see protocol section 6.6.3).

4.5 Additional post-randomisation measures

The number of hospital admissions, as well as the length of stay in hospital is collected. At week 28, the survival data will also be collected from medical notes.

4.6 Timing of measures/participant timeline

A full schedule on the timing of measures is provided in below:

Timepoint		Baseline (≤10 days after screening)	Week≤ 2 (1st rehabilitation intervention)	Week 4 Data collection	Week ≤4 post randomisation (2nd rehabilitation intervention)	Week ≤7 post randomisation (3rd rehabilitation intervention)+	Week 8 Data collection	Week 16 Data collection	Ongoing
Assessment and intervention contact		Face-to-face	Face-to-face	Independent or with investigator	Face-to-face or remote	Face-to-face or remote	Independent or with investigator	Independent or with investigator	
Form:									
1	Informed Consent	x							
2	Registration Form	x							
3	Socio-demographic data	x							
4	Eligibility review	x							
5	Medical History (Comorbidities, clinical diagnosis, treatment history, blood tests results, nutrition and physical activity history)	x							
6	Randomisation	x							
7	AIM, IAM, bespoke questionnaire *		x**				x**		
8	Rehabilitation Data Collection Booklet*		x**		x**	x**			
9	Status form			x			x	x	
10	FACT-G (Primary Outcome)	x		x			x	x	
11	WHODAS 2	x					x	x	
12	IPOS (Physical Symptoms)	x					x	x	
13	Adapted GAS-Light	x					x	x	
14	Modified Client Service Receipt Inventory (CSRI)	x					x	x	
15	Hospital Admissions log								x
16	Adverse Events Log								x
17	Withdrawal form								x

*Participants randomised to the intervention arm only (secondary database)
** Questionnaires/booklets differ for each timepoint
*** Offered to participants who opted-out from 3rd Rehabilitation Intervention visit
+ Session is optional. If participant opts-out then a follow up phone/video call will be offered and documented in the Rehabilitation Data Collection Booklet

Table 4.6: Schedule on the timing of measures

5 Sample Size estimation

5.1 Determination of the primary outcome effect size

Before considering dropout, a sample size of 238 (119 per arm) would provide 90% power at the 2-sided 5% significance level to detect a 5.5-point difference in the mean 8-week FACT-G between arms, adjusting for baseline FACT-G, using analysis of covariance (ANCOVA) or equivalently using the linear mixed effects model planned for the primary analysis. The detectable difference is based on a minimally important difference for FACT-G of 5-6 points derived from multiple approaches and datasets in the relevant population [5].

5.2 Determination of the primary outcome variability

Based on an estimated residual SD of 13.05 from an ANCOVA of the baseline and arm adjusted outcome in the feasibility trial [6], supported by estimated SD of 12.4 for the change from baseline in a relevant trial [7], a standard deviation [8] of 13 was assumed.

5.3 Power to detect effects

There is 90% power to detect the effect using a two-sided 95% confidence interval from an analysis of covariance test with adjustment for baseline FACT-G and randomisation stratifiers.

5.4 Determination of the sample size based on the primary outcome

In order to allow for up to 30% dropout, a sample size of 340 (170 per arm) randomised participants is planned. As the linear mixed effects model makes a missing at random assumption utilising the FACT-G at other timepoints, it is expected that the precision of the estimated intervention effect will be increased in comparison with ANCOVA.
Sample size calculations were performed using nQuery Advisor 4.0 software.

6 Data and Distributions

6.1 Data decisions made

The data manager will make limited decisions about data variables and values so that issues such as missing data can be comprehensively handled by the trial statistician. Decisions which impact on the analysis will be recorded in an appendix of this statistical analysis plan.

6.2 Outcomes requiring derivation

List of outcomes with source of derivation code:

1) FACT-G: As written in the FACT-G scoring manual “subscale scores are calculated by first reversing negatively stated-items (subtracting the response from ‘4’) and then summing the raw (0-4) scores. A total score is then derived by summing subscale scores”[4].

2) WHODAS 2.0: This 36-item questionnaire will be summarised using the simple scoring method, where the scores assigned to each of the items (none, 1; mild, 2;

moderate, 3; severe, 4; and extreme, 5) are summed up without recoding or collapsing response categories) [9, 10].

3) Adapted Goal Attainment Scale – Light measure: A 6-point rating scale is used to record the extent to which a personal goal was achieved (from -2=no change or got worse; 0=as expected; to 2=much better than expected). This information is transformed numerically to produce a single composite GAS t-score for each participant (ranging from 0-100), providing an overall rating or the achievement of goals (maximum 3) for the patients across all the goal areas [11]. The score is computed using the following formula:

   Overall GAS = $50 + \frac{10 \sum w_i x_i}{\sqrt{0.7 \sum w_i^2 + 0.3(\sum w_i)^2}}$, where x_i is the rating of goal *i* and the weight, w_i is taken as one.

For the above three scales, we anticipate these to be approximately normally distributed.

With regards to the IPOS-symptom scale, no derivation is needed, since a summary score (summing the scores of the 10 items) will not be computed but rather the 5-point Likert scale from 0 (not affected) to 4 (overwhelmingly affected) of each symptom will be described.

6.2.1 Procedure for deriving variables

If there is existing syntax code to derive a variable score from questionnaire scoring instructions this will be used . Otherwise new code will be developed by the trial statistician and verified by the senior statistician.

6.2.2 Missing items in scale and subscales

The number (%) of patients with complete data for each scale will be reported. For the scales that provide missing value guidance this will be used. This is the case for FACT-G and WHODAS 2.0 questionnaires.

For FACT-G it is written in the Administration and Scoring Guidelines manual, that if more than 50% of the items comprising a subscale are answered, a subscale score is computed as the prorated sum of the item responses for that subscale. Also, The FACT-G total score is computed as the sum of the four subscale scores, provided the overall item response is at least 80% (i.e. at least 22 of the 27 items were answered) and has a possible range of 0-108 points [4].

For WHODAS 2.0 it is written in the manual: "*Where more than one item is missing:*
- If the respondent is not working and has given responses to the 32-item WHODAS 2.0, the score can be used as it is, and will be comparable to that of the full 36-item version.
- In all other situations where one or two items are missing, the mean score across all items within the domain will be assigned to the missing items. This method should not be used if more than two items are missing. In addition, if domain-wise scores are being computed for domains, the two missing items should not come from the same domain."[9].

6.3 Data transformation

It is not anticipated that any continuous outcomes/ the scales above will need to be considered for transformation, because the scales are bounded and the sample size is reasonably large for group comparisons in the main trial analyses. Assumptions of normality and constant variance required by the models will be examined using residual and other diagnostic plots. If it is relevant, and necessary, such as when sample size is reduced, a log transformation will be considered, because this retains a sensible interpretation for inferences; in relative terms between arms. If an absolute interpretation is needed, then data transformation may not be undertaken, but a nonparametric Bootstrap method for obtaining confidence intervals may be considered [12].

6.4 Defining Outliers

Outliers are observations that have extreme values relative to other observations observed under the same conditions. An outlier will be defined here as a data-point being at least four standard deviations from the mean of its distribution of values observed across patients. This definition will apply to the transformed scale for those outcomes that have been log transformed.
A "bivariate outlier", for the purpose of data checking and querying, will be defined here as a pair of successive serial data-points of the same measure for a participant whose difference is at least four standard deviations from the mean of all patients' such differences. Simple plots of successive pairs of serial measures will be used to assist in identifying outliers for data checking.

6.5 Handling outliers

Outliers will be identified for further investigation by looking at the distributions of the data through histograms, scatter plots or box-plots. Univariate tests for the compatibility of the distribution with a normal distribution will not be undertaken since they can be too sensitive to departures that are often not relevant for the comparison of means (Central Limit Theorem).

Once an outlier is found, a member of the team with sufficient clinical experience will be involved in the decisions as to whether a data value is impossible versus implausible versus plausible. If the outlier is impossible, then it will be set to missing, and a list of these occurrences will be written in the statistical report. If an outlier is clinically plausible, the outlier will remain. If an outlier is clinically implausible (but possible), it will not be ignored or deleted but will be retained for ITT analysis. If outliers remain in the distribution of a variable, then data transformations or nonparametric methods of analysis may be considered.

Sensitivity analysis will be undertaken to check whether the outlier is influential by obtaining results with and then without inclusion of the outlier. If the conclusions are changed, then this will be noted.

7 Descriptive analysis

7.1 CONSORT diagram

A CONSORT diagram of the study, similar to figure in 3.5 above, will be elaborated. It will show the flow of the participants through the stages of the study (eligibility, randomisation, 4, 8 and 16-week follow up), including the number randomised, who comprise the intention to treat population, and the numbers followed-up and to be analysed in primary outcome analysis, as well as the main reasons for non-continuance through the stages of the trial.

7.2 Baseline comparability of randomised groups

Baseline descriptions of participants by arm and overall will be summarised (into Table 1 of the report). No significance testing will be carried out as any differences found may be chance-generated and not for hypothesised reasons.

Continuous variables such as WHODAS 2.0 will be summarised using means and standard deviations [8] and/or medians and interquartile range (IQR, as quartiles Q1 to Q3) for variables presenting a skewed distribution.

Categorical variables such as proportion of patients with pain, in the IPOS-symptom scale, with either 'not at all', 'slight', 'moderate', 'severe' or 'overwhelming/all the time' answers will be described using numbers and percentages, and if ordinal, these may be presented cumulatively numerically and/or graphically.

7.3 Adherence to allocated treatment and treatment fidelity

For the intervention arm to be considered adhered to, the delivery of at least the first two rehabilitation sessions must be completed. The proportions of participants achieving one, two or three sessions, and at least two sessions, will be summarised overall and by arm.

Besides the number of sessions, the following items will also be considered and analysed in the process and implementation evaluation work package (WP6):

1. Duration of the sessions
2. Sessions completed within the required timeframe
3. Location of sessions (hospital/community clinic/participants home/other)
4. Mode of sessions (face to face/remote)
5. Core components received:
    a. Symptom self-management strategies
    b. Strategies to optimise physical activity
    c. Strategies to optimise participation
6. Provided with rehabilitation action plan
7. Provided/recommended for assistive devices/equipment (by equipment type)
8. Provided with material/online self-management resources
9. Signposting/referrals to other services (by category of services referred to)

Descriptive analysis will be used to evaluate if the intervention was delivered with fidelity to the intervention manual as set out in the protocol.

7.4 Loss to follow-up and other missing data

The proportions of participants missing each variable will be summarised in each arm and at each time point.The baseline characteristics of those missing follow-up will be described by arm and overall, and compared to those with complete follow-up.
The reasons for withdrawal from the trial will be summarised overall and by arm.

7.5 Adverse event reporting

Adverse events [7], adverse reactions (AR), serious adverse events (SAE) and serious adverse reactions (SAR) will be summarised overall and by arm.

8 Analysis covariates

8.1 Stratifiers

It is important to consider which, if any, covariates are to be adjusted for in the analyses. The ICH E9 guideline recommends that consideration be given to accounting for randomisation stratifiers by adjusting for them as covariates in linear models. This tends to improve the precision of estimated treatment effects. Therefore, for continuous outcomes, the analysis will include adjustment for the randomisation stratifiers of baseline FACT-G (3 levels), ECOG (2 levels) and country (7 levels).

8.2 Baseline

The corresponding baseline measure for a continuous outcome is also often predictive of the outcome at follow-up. Therefore "baseline" in its continuous linear form will be an additional covariate when modelling continuous outcomes. This will be the case for FACT-G and WHODAS. The continuous baseline will have precedence for inclusion in the model over the corresponding categorical randomisation stratifier, where applicable, such as in the analysis of the primary outcome.

9 Primary outcome analysis

9.1 Statistical Model

The primary outcome analysis approach is informed by guidance on estimands and sensitivity analyses [13].The relevant study objective is to assess the improvement in quality of life from introducing an integrated short-term palliative rehabilitation for those with incurable cancer. The intention to treat [3] population comprises those participants randomised into the trial. They are retained in their randomised arm for the purpose of analysis (irrespective of what they received).

The primary outcome is the FACT-G measured 8-weeks after randomisation. There are additional baseline, 4-week and 16-week FACT-G measurements. The recent ICH-E9 addendum requires an estimand to be defined [14]. The respective attributes are layed out in Table 9.2 below. Principal post-randomisation ("intercurrent") events to consider are the discontinuation of the intervention without having made an action plan in the first meeting, and the death of the participant before an 8-week FACT-G. The population-level summary measure is the absolute difference in population mean 8-week FACT-G between the intervention and comparator arms.

The primary analysis approach will be in the ITT population. A linear mixed effects model (LME) will be fitted, involving the correlated 4-week, 8-week and 16-week FACT-G as the outcomes, allowing different correlations between pairs of FACT-G measurements to be reflected in the model. The covariates will be the arm, the continuous FACT-G at baseline (linear term), ECOG (categorical), and country (categorical). The post-randomisation timepoint (categorical) will also be included as a main effect term and in interaction with each covariate. In the unlikely event of missing baseline covariates, the missing indicator method (White and Thompson 2005) will be used [14]. The removal of the time by country interaction

will be considered if the data is sparse due to drop-out. This model is therefore very similar (in terms of inference and power) to a corresponding Analysis of Covariance (ANCOVA) model but utilises other timepoints to therefore make a more plausible Missing At Random (MAR) assumption than that of the ANCOVA model. An estimate with 95% Confidence Interval (CI) for the 8-week FACT-G intervention effect will be provided, with p-value.

9.2 Estimand

The following table defines our estimand attributes, including the strategies to address the intercurrent events.

Population of interest (common to the 3 timepoints)	All trial participants eligible at baseline		
Variable (endpoint) of interest	FACT-G at 4 weeks	FACT-G at 8 weeks (Primary Outcome)	FACT-G at 16 weeks
Treatment of interest (common to the 3 timepoints)	Up to three rehabilitation intervention sessions with a rehabilitation practitioner		
Intercurrent events:	Strategies for addressing the intercurrent events:		
Death†	Hypothetical	Hypothetical	Hypothetical
Disease deterioration/progression	Treatment policy	Treatment policy	Treatment policy
Palliative rehabilitation discontinuation (Failure to make an Action Plan in the first palliative rehabilitation intervention visit)	Treatment policy	Treatment policy	Treatment policy
Palliative rehabilitation discontinuation due to an AE either related or unrelated to the intervention	Treatment policy	Treatment policy	Treatment policy
Start of another rehabilitation programme*	Treatment policy	Treatment policy	Treatment policy
Population-level summary for the variable:	Difference in FACT-G means between those receiving palliative rehabilitation plus usual care or usual care only at 4 weeks	Difference in FACT-G means between those receiving palliative rehabilitation plus usual care or usual care only at 8 weeks	Difference in FACT-G means between those receiving palliative rehabilitation plus usual care or usual care only at 16 weeks
Analysis method:	LME with additional analyses as described above	LME with additional analyses as described above	LME with additional analyses as described above

**This would be equivalent to a 'rescue medication' therapy in Investigational Medicinal Product trials. Although this can effectively happen, we are not collecting this information since the treatment policy strategy will be considered for the occurrence of this event. This means that the value of FACT-G will be used regardless of whether or not this event occurs.*

†*Antecipated to be very low, since one of the exclusion criteria is 'Clinician rated prognosis of less than 3 months'.*

Table 9.2: Estimands attributes

In summary, the treatment policy strategy (where the value of a subject's FACT-G will used regardless of whether the intercurrent event occurred) will be used for all intercurrent events apart from death, where a hypothetical strategy will be considered. Since there is no FACT-G value if a person dies, we are interested to know what would the FACT-G value be in the 'hypothetical' scenario had the participant not died, making use instead of the linear mixed effects model which will indirectly impute the 'missing' data so that the primary outcome for this participant reflects the quality of life they reported in timepoints while they were alive.

9.3 Sensitivity analysis in respect of missing data

9.3.1 Sensitivity analysis to departures from the MAR assumption

A sensitivity analysis will be undertaken to challenge the plausibility of the MAR assumption in the primary outcome model and tests the overall robustness to missing data, whether or not from study withdrawal or death. This analysis will assess the possibility of alternative plausible values of treatment effect (in size and direction relatively favourable and unfavourable for the intervention) arising from potential mishandling of missing data in the primary analysis model.

For this, we pre-specify a range for FACT-G scores from -10 to +10 over which the mean of the "unobserved outcome data" might *depart* (or be different) from the mean of the "observed outcome data" [15]. In other words, this range can be thought of as how much a typical subject with missing data may on average have had a different estimated treatment effect compared to the corresponding subject with the outcome data observed (given the same baseline covariates and follow-up data in the LME model). The range (-10 to +10) is chosen to represent both negative and positive *departures* that could potentially arise as the "net effect" of alternative reasons which may be unknown; such as dropout due to no anticipated further improvement.

This range of 20 (from -10 to +10) is generously wide for exploring sensitivity of the main results to departures from the MAR assumption, because 10 (as the maximum *departure* in either direction) is larger than the estimated between-arm treatment effect of 5.5 which is a sizeable shift in the mean of the distribution for dropouts compared to completers.

At the end of the trial, the fractions of individuals with missing data for FACT-G at 8 weeks will be available in each arm f_i (for intervention) and f_c (for control). The parameter representing excess FACT-G in those missing compared to those observed, δ, *will take values by passing across the range* -10 to +10. Three scenarios will be undertaken within the sensitivity analysis [16] . These reflect whether departures from the MAR assumption apply within the intervention arm only (rehabilitation in addition to usual care), within the control arm only (usual care), or within both arms equally and in the same direction (thereby potentially cancelling out across the sensitivity range, if the dropout rate were to be the same in both arms).

Scenario 1: the treatment effect from the LME model will be increased by $f_i\delta$
Scenario 2: the treatment effect from the LME model will be increased by $-f_c\delta$
Scenario 3: the treatment effect from the LME model will be increased by $(f_i-f_c)\delta$

This will then provide a series of potential intervention effects with 95% CIs that reveal the degree of robustness to departures from the MAR assumption and allow us to investigate if there is a tipping point at which conclusions would change. The model will also provide the estimated intervention effect and 95% CI for the 4-week and 16-week FACT-G.

9.3.2 Alternative sensitivity analysis to those who died or withdraw from the study due to being too ill to continue

For those participants who died, and those participants who withdrew from the study having mentioned 'too ill' as the main reason for withdrawal, we will increase the δ (departure) to be 50% higher than for the remaining participants, who have other reasons and/or were lost to follow-up. In other words, we will have an f_1 and an f_2 for each of the arms, where f_1 would represent the fraction who died or dropped out because of being too ill and f_2 the fraction of those who dropped out for other or unknown reasons.

9.4 Sensitivity analysis to adherence to the intervention

The intercurrent event of not complying to produce an action plan is ignored in the primary outcome analysis approach above due to the emphasis on the ITT approach which takes a treatment policy strategy for this event and accepts all FACT-G outcomes into the analysis to answer the scientific question, pragmatically accepting the influence on these from reduced intervention compliance.

Given that those who do not complete an action plan and so do not attend the first intervention session (which should occur as close to randomisation as possible and no later than 14 days after), may have a more similar experience with those in the control arm, the difference in outcomes maybe be smaller than anticipated. Therefore, we will conduct an analysis estimating the effect of the rehabilitation versus usual care group on the primary outcome in a more highly compliant population, and restricted to those confirmed eligible, whilst respecting randomisation. This approach should provide a better estimate of the true effect of intervention with ideal compliance, and without suffering from potential biases seen in a per-protocol analysis.

Therefore a complier average causal effect (CACE) analysis will be carried out as recommended and outlined by Dunn et al.[17]. This estimate is the comparison of the average outcome of the compliers in the rehabilitation arm with the average outcome of the comparable group ("would-be compliers") in the usual-care arm.

The outline of the approach to be taken is given here:
Sample sizes (N) and means (M) are deduced for the Standard Care control "would-be compliers" and "would-be-noncompliers" in the following table by assuming that the proportion of intervention group compliers, and control group would-be compliers, is the same under randomisation, and that would-be non-compliers in the control group would have the same mean outcome as non-compliers in the intervention group (the exclusion restriction assumption). The sample sizes refer to those followed-up with primary outcome data (FACIT-G at 8 weeks).

Arm	Compliers (making an action plan)	Noncompliers (not making an action plan)	All
Rehabilitation arm	N_{I1} M_{I1}	N_{I2} M_{I2}	N_I M_I
Usual-care arm	$= N_C - (N_{I2}/N_I)*N_C$ $= (M_C-(N_{I2}/N_I)*M_{I2})/(N_{I1}/N_I)$	$= (N_{I2}/N_I)*N_C$ $= M_{I2}$	N_C M_C

(Statistics preceded by an "=" are unobserved, and are estimated from the observed statistics.)
The method is adapted to a more plausible MAR assumption by replacing the sample sizes at follow-up by those at baseline. In the presence of missing compliance, it will be primarily assumed that the participant is a non-complier.

The CACE estimate will instead be obtained from the primary analysis LME model. It is the ratio of the estimated treatment effect to the proportion compliant, following the rule of thumb (estimate LME/ proportion compliant)[18]. 95% confidence intervals for this estimate will be provided.

A secondary CACE estimate will be obtained by considering compliance as those who attended at least the first two intervention sessions. However, this is will be given less importance since it is unlikely that the mean outcome of the non-compliers in the intervention group is similar to the mean outcome of the would-be non-compliers in the control group as per assumption of CACE.

9.5 Supplementary analysis of the estimand

Previous sensitivitiy analysis already examines the overall robustness to missing data, whether or not from study withdrawal or death. This supplementary analyses will add additional insights into the understanding of the intervention effect.

The intercurrent event of death is important to consider in palliative and end-of-life trials [19]. Deaths are expected to be relatively rare in the initial weeks after randomisation due to the eligibility criteria excluding those with short prognosis, and rarer still within the ECOG 2 stratum (covered in the subgroup analysis section), and the intervention is not expected to affect timing of death.

Nevertheless, death of a participant in either arm may occur before the primary outcome is provided at its intended collection point by the 8-week + 3-day window point.

In this scenario, for those participants that died and provided a 4-week FACT-G assessment, the following analysis will be considered:

(i) the 4-week FACT-G is taken to fully represent such a participant's end-of-life period, and is replaced as the 8-week outcome and/or

(ii) those who have died are prevented from the primary outcome model's implicit 8-week FACT-G imputation after death, by removal of the 4-week outcome data.

These contrast with the primary analysis, which implicitly imputes FACT-G after the death of a participant for any cause, just as it will after withdrawal or other missing data in the follow-up timepoints, reflecting a "hypothetical strategy".

9.6 Planned subgroup analysis

The following subgroup variables will be considered so as to investigate the differences between groups in each of the categories of the variables

i) Gender (Woman, Man, Other, Prefer not to answer)
ii) Age (<65; ≥65)
iii) Diagnosis (locally advanced or metastatic disease)
iv) ECOG performance status (2, 3)
v) Country (7 countries currently)
vi) Living situation (alone, with others)
vii) Dependents with care needs (yes, no)
viii) Health confidence (high, low)

With regards to age, we hypothesize that there may be a potential benefit of the intervention in those age 65 or older, and the same for higher disability (diagnosis and higher ECOG status)[20].

The consistency of the primary outcome result will be examined across categories of subgroup variables. The LME model will be extended to include 2-way interactions between trial arm and the subgroup variables. Effect sizes with 95% confidence intervals for each prognostic group will be estimated.

These analyses have relatively high variability to be able to make statistically robust conclusions, therefore caution will be exercised in the reporting and interpretation of the estimates and 95% confidence intervals obtained from these analyses.

9.7 Interim analysis

Formal interim analysis of the primary outcome for early stopping is not planned for this study. Regular interim reports will be prepared as needed for DMC meetings.

10 Secondary outcome analysis

10.1 Analysis of continuous outcomes

As for the primary outcome, the analysis of continuous secondary outcomes will be each involve comparison between arms using a linear mixed effect model adjusting for the minimisation stratifiers, baseline FACT-G (3 levels), ECOG (2 levels) and country (7 levels currently) and the baseline of the outcome with the associated missing indicator. Time will be represented as categorical contrasts in main effect form and in interaction with all other fixed effects.

10.2 Analysis of binary outcomes

For the binary outcomes, such as for estimating the prevalence of adverse events, summaries will be reported as unadjusted patient proportions and rates within and between arms with 95% confidence intervals using exact methods where appropriate.

10.3 Summary of the analysis methods for secondary outcomes

It is planned that all study analyses will use methods that provide two-sided 95% confidence intervals.
For the secondary outcomes mentioned in section 6.2, the following analysis will be used:

Outcomes:	Analysis method:
FACT-G at (28 5-point items) 4 and 16 weeks Domain subscales: Physical well-being (0- 28) Social/family well-being (0- 28) Emotional well-being (0- 24) Functional well-being (0- 28) Total score (0- 108 best quality of life)	Adjusted difference in means using linear mixed effects model

WHODAS 2.0 (36 5-point items) at 8 and 16 weeks Domain subscales: Undestanding and communicating (6- 30) Mobility (0- 25) Self-care (4- 20) Getting along interacting with other people (5- 25) Life activities (4-20 or 8-40 depending on working status) Participation (8- 40) Total score (36 -180 extreme difficulty)	Adjusted difference in means using linear mixed effects model
IPOS- symptoms at 8 and 16 weeks Symptoms (5-point scale): Pain Shortness of breath Weakness or lack of energy Nausea Vomiting (begin sick) Poor appetite Constipation Sore or dry mouth Drowsiness Poor mobility	Proportions in each category will be summarised. The AUC with 95% confidence intervals at each timepoint will be performed*. Equivalent to non-parametric Mann-Whitney U test analysis.
Goal attainment score (0- 100). Each goal rated on a 5-point scale	Adjusted difference in means using linear mixed effects model
Prevalence of adverse events	Differences in proportions with 95% confidence intervals

*The change from baseline may also be considered.

11 Handling multiple comparisons

Significance tests will be used sparingly and restricted where possible to addressing stated hypotheses. Secondary outcomes, as well as the primary outcome, will be summarised using an effect size with a 95% confidence interval. Interpretation for those secondary outcomes that do not directly address the stated study hypotheses will be more cautious.

12 Software

Data management:
An online data collection system for clinical trials (MACRO; InferMed Ltd) will be used. This is hosted on a dedicated server at KCL and managed by the Mental Health and Neuroscience Clinical Trials Unit (CTU) at the Institute of Psychiatry in London . The CTU Data Manager will extract data periodically as needed and provide these in comma sepa (.csv) format.

Statistical analysis:
The principal software package will be IBM SPSS Statistics 28 and R software will be available.

13 DMC monitoring

We expect the DMC would want to monitor the the investigational arm in relation to the standard arm and we would regularly provide information such as participant adherence to rehabilitation sessions, withdrawals and other information.

14 Acknowledgments

In translating the study protocol into this statistical analysis plan, we are grateful to explanations from the study team including Joanne Bayley and Matthew Maddocks. Further versions of the plan will be commented on by members of the Data Monitoring and Trial Steering Committees.

15 Amendments to Versions

Version 0.1 was written by Joana Vasconcelos on 17th May 2024 and revised by the Senior statistician on 21th May 2024. Version 0.2 was completed on 18th June 2024 after revision of the chief investigator on 3rd July 2024. Version 1.0 was completed on 30th October 2024 after revision of the TSC.

Amendments to versions of the SAP will be listed here.

Reference list

1. Harmonisation, I.C.f.; Available from: https://www.ich.org/page/efficacy-guidelines.
2. Schulz, K.F., D.G. Altman, and D. Moher, *CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials.* Journal of Pharmacology and pharmacotherapeutics, 2010. **1**(2): p. 100-107.
3. Products, C.f.P.M., *Committee for proprietary medicinal products (CPMP) points to consider on adjustment for baseline covariates.* Statistics in medicine, 2004. **23**(5): p. 701-709.
4. *Functional Assessment of Cancer Therapy - General*. [cited 2024 05]; Available from: https://www.facit.org/measures/fact-g.
5. Eton, D.T., et al., *A combination of distribution-and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale.* Journal of clinical epidemiology, 2004. **57**(9): p. 898-910.
6. Bayly, J., et al., *Short-term integrated rehabilitation for people with newly diagnosed thoracic cancer: a multi-centre randomized controlled feasibility trial.* Clinical rehabilitation, 2020. **34**(2): p. 205-219.
7. Quist, M., et al., *Effects of an exercise intervention for patients with advanced inoperable lung cancer undergoing chemotherapy: a randomized clinical trial.* Lung Cancer, 2020. **145**: p. 76-82.
8. Corrie, P.G., et al., *Is community treatment best? a randomised trial comparing delivery of cancer treatment in the hospital, home and GP surgery.* Br J Cancer, 2013. **109**(6): p. 1549-55.
9. Üstün, T.B., *Measuring health and disability: Manual for WHO disability assessment schedule WHODAS 2.0*. 2010: World Health Organization.
10. *World Health Organization Disability Assessment Schedule 2.0*. [cited 2024; Available from: https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health/who-disability-assessment-schedule.
11. Turner-Stokes, L., *Goal attainment scaling (GAS) in rehabilitation: a practical guide.* Clinical rehabilitation, 2009. **23**(4): p. 362-370.
12. Carpenter, J. and J. Bithell, *Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians.* Statistics in medicine, 2000. **19**(9): p. 1141-1164.
13. *Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principals for clinical trials ICH E9(R1)*. Available from: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.
14. White, I.R. and S.G. Thompson, *Adjusting for partially missing baseline measurements in randomized trials.* Statistics in medicine, 2005. **24**(7): p. 993-1007.
15. White, I.R., J. Carpenter, and N.J. Horton, *Including all individuals is not enough: lessons for intention-to-treat analysis.* Clinical trials, 2012. **9**(4): p. 396-407.
16. White, I.R., E. Kalaitzaki, and S.G. Thompson, *Allowing for missing outcome data and incomplete uptake of randomised interventions, with application to an Internet-based alcohol trial.* Statistics in medicine, 2011. **30**(27): p. 3192-3207.
17. Dunn, G., M. Maracy, and B. Tomenson, *Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods.* Statistical methods in medical research, 2005. **14**(4): p. 369-395.
18. White, I.R., *Uses and limitations of randomization-based efficacy estimators.* Statistical Methods in Medical Research, 2005. **14**(4): p. 327-347.

19. Hussain, J.A., et al., *Missing data in randomized controlled trials testing palliative interventions pose a significant risk of bias and loss of power: a systematic review and meta-analyses.* Journal of clinical epidemiology, 2016. **74**: p. 57-65.
20. Cao, C., L. Yang, K.H. Schmitz, and J.A. Ligibel, *Prevalence and cancer-specific patterns of functional disability among US cancer survivors, 2017-2022.* Journal of Clinical Oncology, 2024: p. JCO. 23.02536.