Multidisciplinary home-based rehabilitation in inoperable lung cancer – a multi-site randomised controlled trial

Statistical analysis plan

Introduction

This document described the plan for the statistical analysis of a phase II randomized controlled trial examining the efficacy of an intervention of home-based multidisciplinary rehabilitation in non-small cell lung cancer (NSCLC).

Inclusion criteria

To be eligible participants must:

- have a diagnosis of inoperable NSCLC,
- be scheduled to receive treatment for the primary lung tumour other than surgery (i.e. chemotherapy, radiotherapy or targeted therapy),
- have commenced treatment no more than four weeks prior to recruitment,
- be aged 18 years or older,
- be able to read and write English,
- have an Eastern Co-operative Oncology Group (ECOG) performance status of less than or equal to two and a Clinical Frailty Scale (CFS) score of less than seven;
- have a physician rated life expectancy greater than six months and the treating oncologist's approval for study involvement.

Exclusion criteria

Participants are excluded if they:

- have a concurrent, actively treated other malignancy (or one-year history of other malignancy (three-years for breast cancer due to proximity of the radiotherapy treatment field to the lung)) other than non-melanoma skin cancer or in-situ melanoma,
- have any co-morbidities or evidence of pelvic or lower limb bony metastases prohibiting participation in a land-based exercise program,
- met physical activity guidelines in the past month based on self-report (150 minutes or more of moderate intensity physical activity per week),
- have a current unstable psychiatric or cognitive disorder.

Sources of study participants

Study participants will be recruited from three hospitals in Melbourne where eligible NSCLC patients are cared for. The hospitals form part of the Victorian Comprehensive Cancer Centre Alliance, and are The Peter MacCallum Cancer Centre, the Royal Melbourne Hospital and The Austin Hospital.

Trial aims

The **primary** aim of the trial is to assess the efficacy of targeted home-based, multi-disciplinary exercise and supportive care (hereafter, the "intervention"), on the change from baseline to nine weeks in functional exercise capacity (6 Minute Walk Distance, 6MWD) in people with inoperable NSCLC.

Secondary aims are as follows:

To assess the efficacy of the intervention on physical activity levels, muscle strength and patient-reported outcomes, in terms of the change from baseline to nine weeks;

To assess the efficacy of these outcomes in terms of the change between baseline and follow-up measurements at 24 weeks;

To compare the overall survival, and progression free survival, between the intervention group and the usual care group.

Exploratory aims

In a convenience sub-sample of study participants, ultrasound measurements at baseline and 9 weeks will be obtained and changes will be analysed by study group.

In a different sub-sample of study participants, venous blood measurements at baseline and 9 weeks will be obtained and changes will be analysed by study group.

Primary outcome

The primary outcome variable is the *change* between 6 Minute Walk Distance test (6MWD) values evaluated at 9 weeks post-baseline assessment, and at baseline.

Schedule of the measurement of outcomes

The following table shows the outcomes that will be measured, and the time-points at which they will be recorded. The abbreviations used in this table are expanded in the material that follows.

Outcomes	Time point			
	Baseline	**Post-program (9-weeks)**	**4-months (telephone)**	**6-months**
Primary outcome				
6MWD	✓	✓		✓
Key secondary outcomes				
Physical Activity				
Accelerometry	✓	✓		✓
IPAQ	✓	✓		✓
Strength				
HHD quadriceps	✓	✓		✓
HGD	✓	✓		✓
Secondary outcomes				
HRQoL				
FACT-L	✓	✓		✓
AQoL	✓	✓		✓
PAAI	✓	✓		✓
BREQ-2	✓	✓		✓
MDASI-LC	✓	✓		✓
HADS	✓	✓		✓
CD-RISC	✓	✓		✓
Health economic questionnaire		✓	✓	✓

Qualitative interviews (subset)		✓
Exploratory outcomes (subset)		
Venous blood sample	✓	✓
Quadriceps size and echogenicity	✓	✓

Measurement protocols

The protocols for measuring outcomes are as follows.

Six Minute Walk Distance test (6MWD)

On each occasion when the 6MWD test is carried out, two separate distances are recorded in the standard protocol of this test. The larger of the two distances walked (m) of the two tests performed will be used in analyses.

Accelerometry

Objective motion sensors (Sensewear™ armbands) will be used. These are worn on the posterior aspect of the participant's upper arm for a period of seven days. The daily wear time will be reported. Averages per day will be recorded for the following measures: steps, energy expenditure, metabolic equivalent of tasks (METs), sedentary time and time spent in light, moderate and vigorous physical activity.

The minimum data requirement is four days of eight hours monitoring; otherwise, these measures will be regarded as missing.

International Physical Activity Questionnaire (IPAQ), Short Form

The scoring for this will follow the algorithms described in the "Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ) - Short and Long Forms (November 2005)". The variables derived from the questionnaire are a continuous score, measured in MET-minutes (MET = Metabolic Equivalent of Task, or metabolic equivalent), and a categorical score, measured as one of three levels, "low", "medium" or "high". The items making up the scores ask for a number of days during the last 7 days that a defined level of activity was carried out, and the amount of time usually spent doing the activity. The responses for the amount of time usually spent on one of those days have 'don't know/not sure' as an option. If this response has been used, or if data are otherwise missing for time or days, the participant's response will be taken as missing for this outcome.

For available data, comparisons will be made between patient-reported (IPAQ) and objectively-measured (Sensewear accelerometers) physical activity at baseline, 9 weeks and 24 weeks.

Hand-held dynamometer (HHD) quadricep strength

An isometric maximal voluntary contraction will be assessed and the highest force achieved over five seconds will be recorded, three times. The highest of the three tests, and highest of left of right sides, will be used for analyses and will be reported in Newton meters (Nm).

Hand-grip dynamometer (HGD) grip strength

Three consecutive efforts are made with five seconds of rest in between each repetition on both sides. All three efforts are recorded and reported in kilograms (kg). The highest value of the three tests and left or right sides will be used in analyses.

Functional Assessment of Cancer Therapy-Lung (FACT-L)

The FACT-L is a 36-item questionnaire containing nine lung cancer specific questions. The Trial Outcome Index (FACT-TOI) is a measure that sums the functional wellbeing (FWB), physical wellbeing (PWB), and the lung cancer (LCS) subscales of the FACT-L.

FACT-TOI is scored by summing the individual scale scores, with higher scores indicating better quality of life.[1] Each domain, as well as the overall QOL score is calculated according to the scoring instructions for FACT. After reversing the scoring of negatively worded items (so that a higher score always indicated a favourable response), item responses are summed. The average value of the items for a subscale is computed for missing values, as long as more than 50% (hence, 4 out of 6, or 4 out of 7) of the questions in the subscale were answered.

The FACT-L scale is considered to be an acceptable indicator of patient quality of life if:

- the overall item response rate is greater than 80%.
- all of the component subscales have valid scores.

The same rules apply to FACT-TOI.

Assessment of Quality of Life, Version 2 (AQoL v1)

AQoL version 1 is an overall measure of health-related quality of life. The scoring is described in by Hawthorne et al.[2]

When a value is missing from one variable within a dimension, its value is imputed to be the mean integer value of the two other variables within that dimension. For example, if the data were: Q10: X, Q11: 1, Q12: 3, the program imputes the value for Q10 as (1+3)/2 = 2.

Where intermittent missing data comprise 30% or more of all responses the data will be regarded as missing.

Physical Activity Assessment Inventory (PAAI)

This is a single score derived from 13 items scored on a 0-100 point scale. The average of the 13 items is used.

If three or more of the 13 items are missing, the PAAI will be regarded as missing. If one or two of the 13 items are missing, the score will be calculated as the average of the available items.[3]

Behavioural Regulation in Exercise Questionnaire (BREQ-2)

BREQ-2 is a 19-item tool consisting of five subscales measuring a continuum of behavioural regulation in exercise. The subscales are: external (4 items), introjected (3 items), identified (4 items),

intrinsic (4 items) and amotivation (4 items). An overall score of participant self-determination is derived from the sum of the subscale average scores. If two or more items of a subscale are missing the subscale will be taken to be missing. If one item of a subscale is missing the average score of the answered items will be used as the subscale score.

MD Anderson Symptom Inventory – Lung Cancer (MDASI-LC)

The MDASI-LC score is based on an inventory of symptom severities (19 items, 13 "core", 3 lung cancer specific items and 6 "interference".[4,5] Mean scores for symptom severity (13 core and 3 lung cancer specific items), core (13 items) and interference (6 items) are reported.

The MDASI assesses the severity of symptoms at their worst in the last 24 hours on a 0–10 numeric rating scale (NRS), with 0 being "not present" and 10 being "as bad as you can imagine." A component score for the MDASI symptom severity scale is obtained by taking the average of the 13 core items together. When more than 50% of the items (7 out of 13) are recorded, the scale can be calculated as the average of the available items; otherwise, the scale is regarded as missing.

For the 3 lung cancer items, if more than 1 of the 3 items is missing the lung cancer specific scale will be regarded as missing, otherwise the average of the other 2 responses will be used.

The mean of the 6 "interference" items can be used to report on symptom distress. For this mean to be used a minimum of 4 of the 6 items need to be completed (sum of items answered) x 6 / (number of items answered).

The three lung-specific symptom items are not for use individually but can be reported with the 13 MDASI core symptom items (for example, examining the relationship between a MDASI core symptom and a lung-specific symptom item which both score highly). Individual or subsets of symptoms can also be reported, if defined a priori, in the lung cancer population drowsiness, fatigue, sleep disturbance, shortness of breath, and pain are likely to be items which score more severely.

Hospital Anxiety and Depression Scale (HADS)

The 14-item Hospital Anxiety and Depression Scale (HADS) is used to screen for anxiety and depression symptoms. When at least 50% of the items (7 out of 14) are recorded, the scale will be calculated as the average of the available items; otherwise, the scale will be regarded as missing.[6]

Connor Davidson Resilience Scale, 10 items (CD-RISC-10)

The CD-RISC-10 is a measure of resilience based on a 10-item scale. The scale is obtained by summing the total of all items, each of which is scored from 0-4. The score range for the 10-item scale is therefore from 0-40, with higher scores reflecting greater resilience.

In order to consider the scale to be valid, at least 70% (7 items of the CD-RISC-10) should have been completed. When there are one, two or three items are missing, they will be assigned the average of the available items, and the overall score will then be calculated. If six or fewer items are available, the score will be regarded as missing.

Other outcomes

The ultrasound variables measured, if feasible, will include rectus femoris cross-sectional area and thickness, and echogenicity.

The venous blood variables measured, if feasible, will include interleukin-6 (IL6) and C-reactive protein (CRP).

All-cause mortality

Mortality will be recorded up until three years follow-up; where patients withdraw or become lost to follow-up the last known date of contact will be recorded.

Progression-free survival

Progression is defined to have occurred on the first date that diagnostic imaging (CT or PET scan) is reported as showing progression of disease. For patients who die due to disease progression prior to any re-staging imaging date of progression will be the same as the date of death. If progression has not occurred by three years of follow-up or by the date at which the subject withdraws or is lost to follow-up, the outcome will be censored at that date.

Analysis principles and methods

All analyses will be conducted on an intention-to-treat basis to the extent possible (see "Dealing with missing values", below), with subjects retained in their original assigned groups, and will be unadjusted for the effects of other covariates except where indicated. Where there are missing observations, the number of observations used in the analysis will be reported. Complete case analysis (listwise deletion) will be undertaken for background comparison to the reported multiple imputation analyses.

The differences between the two groups at baseline will be described in a standard way, with percentages for categorical variables and means and standard deviations for continuous variables. No hypothesis tests will be carried out for these comparisons.

Estimates and 95% confidence intervals will be reported. The key emphasis in reporting results will be on directly interpretable treatment effects comparing the two study groups, such as differences of means, differences of proportions, and hazard ratios, at 9 weeks follow-up and 24 weeks follow-up.

Analysis of outcomes

The primary outcome, the change in the 6MWD from baseline to 9 weeks, will be compared between the two groups, using a standard t-test, with transformations to support the necessary assumptions of the analysis, if necessary. All secondary outcomes will be assessed in this manner comparing the change from baseline to 9 weeks and baseline to 6 months between the two groups, to make use of all available data.

Analyses of continuous outcomes

Many of the outcomes are continuous scores measured at baseline, 9 weeks and 6 months. For such outcomes (e.g. HADS), the 9-week and 6-month results will also be analysed with adjustment for baseline, using a linear model with study group and baseline outcome as explanatory variables. This will produce an estimate and confidence interval for the treatment effect at 9 weeks, and also at 6 months. If necessary, transformations will be used to support the necessary assumptions of the analysis.

Distribution-free approaches

Prior to examination of the data, it is not possible to say whether all of the continuous variables can be accommodated in a linear model framework, even with the use of transformations. If necessary, distribution-free approaches will be used to make relevant comparisons at the two follow-up times. These include the Mann-Whitney test.

Analysis of IPAQ, categorical scale

The data arising from IPAQ has a categorical scale which has the values "low", "medium" or "high" levels of physical activity. It is not expected that there will be many observations of "high" physical activity, given the condition being studied, and the criteria required to satisfy the definition of this level. Accordingly, this variable will be collapsed, for analysis purposes, to two categories, "low" and ("medium or high"). This derived variable will be analysed using logistic regression with random effects, analogous to the analysis of the continuous outcomes.

Per protocol analysis

After the blinded intention to treat analysis is completed, the statistician will be advised of the adhering subjects in the intervention group, and an analysis will be carried out of the compliant group versus the control group (omitting the non-adhering subjects in the intervention group). "Adhering" will be defined as completing at least two aerobic sessions per week for six of the eight-week initial program. The outcomes to be analysed are 6MWD, steps per day, moderate-vigorous physical activity (minutes/day), IPAQ continuous (MET-minutes per week) and categorical scores, quadriceps and hand-grip strength.

Per protocol analysis of survival

To be considered as satisfying the protocol of the intervention, and hence to be considered "adherent", as defined in the above paragraph, participants have to survive at least 8 weeks post-baseline. Therefore, to carry out the per protocol analysis of survival, the same restriction will be imposed on the usual care group, to allow an unbiased comparison.

Subgroup analyses

Analyses of 6MWD will be carried out for the following subgroups:

Baseline 6MWD, divided into tertiles

Steps per day, divided into tertiles

For treatment intent: radical or palliative (two groups) subgroup analyses will be undertaken for 6MWD, health-related quality of life (AQoL and FACT-L) and symptom severity and interference.

Dealing with missing values

It is expected that there will be missing data arising from a variety of sources. The description of the measures (above) deals with the cases where a missing observation may arise from partial responses on a scale, where a subject has provided responses for some but not all items.

In addition, there will be data missing because a study participant does not provide data at one or both of the follow-up times, 9 weeks and 24 weeks.

The reasons for this could be withdrawal, loss to follow-up, or death.

It is also possible that, at a follow-up time, a subject may provide some measures but not others; for example, they may provide self-reported data (HADS, etc.) but be unwilling to provide the 6MWD test.

Despite a selection criterion being that all subjects must have a physician-rated life expectancy of more than 6 months at enrolment, it is expected there will be deaths in both study groups before the first follow-up, and also between the 9-week and 24-week follow-ups.

When the death occurs very early, it is even possible that the subject will not have experienced any treatment, if they are allocated to the treatment group.

Provided that some data are available at 9 weeks, multiple imputation will be used to impute missing data at 9 weeks, and missing data at 24 weeks, with one exception. When a subject dies between 9 and 24 weeks, no data will be imputed for the 24-week follow-up. It is widely accepted, and accords with common sense, that the imputation of missing data on a health-related outcome (HADS etc.) for someone who has already died should not be done.

It is considered that the number of early deaths will be small, and that the early death rates in the two study groups will not be markedly different. Therefore, no attempt will be made to adjust for mortality in the analysis of the other health-related outcomes.

To deal with the issue of the omission of deaths and subjects with no data post baseline, in the analysis of the primary outcome only, a distribution-free sensitivity analysis of a composite outcome for all 92 subjects at follow-up will be undertaken. In this analysis subjects who have died will be considered to have a worse outcome that anyone surviving, whose 6MWD is recorded. Rather than being assigned a score of zero, these subjects will be assigned a value lower than the minimum observed 6MWD because the value allocated affects the linked equations in the multiple imputation, and therefore needs to be not extreme. Using the Mann-Whitney test, a distribution free test based on ranks only, means that the deaths can be counted as having a value lower than any observed value. Given that the ranks depend only on the order of the data, this does not amount to any numerical assumption about the value used.

For the subjects without a follow-up 6MWD, but not known to be dead, multiple imputation, involving 20 multiply imputed data sets, will be used. A composite outcome will be analysed for all 92 subjects, using the Mann-Whitney test, to test the difference between the intervention and usual care groups for the 6MWD. The observed 6MWD values will be used, deaths at follow-up will be assigned a value less than the minimum observed value, and living subjects with missing data will be included in analyses using multiple imputation.

Survival analysis

Analysis of all-cause mortality will be carried out on a strict intention to treat basis (all patients in their randomised groups will be included). The analysis will include comparisons between the two study groups using Kaplan-Meier curves and the log-rank test. Median survival will be estimated and compared using a 95% confidence interval. The estimate of the hazard ratio for survival, with 95% confidence interval and P-value, will be obtained from Cox's proportional hazards model.

Analysis of progression-free survival will follow the same procedure.

Secondary analysis of all-cause mortality will be carried out on a per protocol basis, comparing the adherent patients in the intervention group with the usual care group.

References

1. Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. *Health & Quality of Life Outcomes* 2003; **1**: 1-7.
2. Hawthorne G, Richardson J, Osborne R. The Assessment of Quality of Life (AQoL) instrument: A psychometric measure of health-related quality of life. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation* 1999; **8**(3): 209-24.
3. Haas B. Fatigue, Self-efficacy, Physical Activity, and Quality of Life in Women With Breast Cancer. *Cancer nursing* 2011; **34**(4): 322-34.
4. Cleeland CS. The M.D. Anderson Symptom Inventory. User Guide. Version 1: The University of Texas MD Anderson Cancer Centre, 2014.
5. Mendoza TR, Wang XS, Lu C, et al. Measuring the symptom burden of lung cancer: the validity and utility of the lung cancer module of the M. D. Anderson Symptom Inventory. *The oncologist* 2011; **16**(2): 217-27.
6. Bell ML, Fairclough DL, Fiero MH, Butow PN. Handling missing items in the Hospital Anxiety and Depression Scale (HADS): a simulation study. *BMC research notes* 2016; **9**(1): 479.