  

**Statistical Analysis Plan**

**Version 1.1**

**02nd February 2017**

Study Title **Bridging the Age Gap in Breast Cancer: Improving outcomes for older women. (Cohort study)**

Funding body **National Institute for Health Research (NIHR) Programme Grant (RP-PG-1209-10071)**

Sponsor **Sheffield Teaching Hospitals NHS Foundation Trust**

Authored by

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_/\_\_/\_\_

*Oscar Bortolami*

*Statistician*

*ScHARR. University of Sheffield*

Approved by

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_/\_\_/\_\_

*Stephen Walters*

*Professor of Medical Statistics and Clinical Trials*

*ScHARR, University of Sheffield*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_10\_/10\_/16\_\_

*Lynda Wyld*

*Reader in Surgical Oncology*

*Department of Oncology and Metabolism, University of Sheffield Medical School*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_/\_\_/\_\_

*Sue Ward*

*………..*

*ScHARR, University of Sheffield*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_/\_\_/\_\_

*Alistair Thompson*

*……..*

*DMEC chair*

SAP HISTORY

|  |  |  |  |
| --- | --- | --- | --- |
| **Version Date** | **Date**  **approved** | **Modifications (with section)** | **Prior to/After blind review**  **Prior to/after database lock** |
| 1.1 |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Table of Contents**

[1 Introduction, study design and key trial objectives 5](#_Toc464462335)

[1.1 Study outline 5](#_Toc464462336)

[1.2 Outcome measures 6](#_Toc464462337)

[1.2.1 Primary outcome 6](#_Toc464462338)

[1.2.2 QOL related secondary outcomes 6](#_Toc464462339)

[1.2.3 Survival related secondary outcomes 7](#_Toc464462340)

[1.2.4 Other secondary outcomes 7](#_Toc464462341)

[1.3 Sample size 7](#_Toc464462342)

[1.3.1 PET versus surgery analysis 7](#_Toc464462343)

[1.3.2 Chemotherapy analysis 8](#_Toc464462344)

[1.4 Data monitoring 8](#_Toc464462345)

[1.4.1 Data Monitoring and Ethics Committee (DMEC) 8](#_Toc464462346)

[1.4.2 Trial Management Group (TMG) 8](#_Toc464462347)

[1.4.3 Interim analysis 8](#_Toc464462348)

[2 Data sources, protocol non-compliance and analysis populations 9](#_Toc464462349)

[2.1 Data sources 9](#_Toc464462350)

[2.2 Protocol non-compliances 9](#_Toc464462351)

[2.3 Analysis populations 9](#_Toc464462352)

[3 Outline of statistical analyses 9](#_Toc464462353)

[3.1 General considerations 9](#_Toc464462354)

[3.2 Demographics and baseline characteristics 9](#_Toc464462355)

[3.2.1 Definitions and data manipulation 10](#_Toc464462356)

[3.2.2 Recruitment and data completeness 14](#_Toc464462357)

[3.3 Clinical outcomes 15](#_Toc464462358)

[3.3.1 QOL key secondary outcome 15](#_Toc464462359)

[3.3.2 Statistical analysis for the QOL key secondary outcome 15](#_Toc464462360)

[3.3.3 Quality of life analysis 17](#_Toc464462361)

[3.3.4 Analysis of survival related outcomes 17](#_Toc464462362)

[3.3.5 Subgroup evaluation 18](#_Toc464462363)

[3.3.6 Model diagnostics 18](#_Toc464462364)

[3.4 Safety outcomes 20](#_Toc464462365)

[3.5 Missing and spurious data 20](#_Toc464462366)

[3.6 Health economic analyses 20](#_Toc464462367)

[4 Appendix 22](#_Toc464462368)

[4.1 Tables Output 22](#_Toc464462369)

[5 References 42](#_Toc464462370)

**List of abbreviations**

ADL Activities of daily living

AE Adverse event

AFT Accelerated failure time

ATT Average treatment effect of the treated

BCCOM Breast Cancer Clinical Outcome Measures

CI Chief Investigator or confidence interval

CRF Case Report Form

CTCAE Common terminology criteria for adverse events

CTRU Clinical Trials Research Unit

DMEC Data Monitoring and Ethics Committee

DSI Decision Support Instrument

ECOG PS Eastern Cooperative Oncology Group Performance Status

EORTC European Organisation for Research and the Treatment of Cancer

EQ-5D EuroQol 5D

ER Oestrogen receptor

FFS Failure-free survival

GCP Good Clinical Practice

GEE Generalised estimating equation

HES Hospital Episode Statistics

HR Hazard ratio

IADL Instrumental activities of daily living

ICH International Conference on Harmonisation

MMSE Mini-mental state examination

OS Overall survival

PET Primary Endocrine Therapy

PS Performance Status

QALY Quality Adjusted Life Year

QoL Quality of life

SAE Serious adverse event

SAP Statistical analysis plan

SDV Source data verification

SOP Standard operating procedure

TMG Trial Management Group

# 1 Introduction, study design and key trial objectives

## 1.1 Study outline

The Bridging the Age Gap study is a non-randomised, pragmatic, cohort study designed to observe normal UK clinical practice for the treatment of older women with breast cancer.

The objectives of the study are as follows:

Primary objective: development of a prognostic model in patients undergoing either Primary Endocrine Therapy (PET) or Chemotherapy

PET

1. To determine the patient and cancer characteristics which predict whether PET is a safe and effective breast cancer treatment in older women with oestrogen receptor positive (ER+) breast cancer by means of statistical modelling based on both retrospective registry data, hospital episode statistics (HES) data and prospective cohort study data.
2. To develop a simple scoring system, based on co-morbidity, dependency, age and tumour characteristics, which will enable prediction of those women best treated with PET or surgery.
3. To develop a web-based algorithm based on the developed model to aid clinician decision making.

Chemotherapy

1. To determine the patient and disease characteristics that predict whether the addition of post-operative adjuvant chemotherapy results in improved outcomes in older women with high prognostic risk, operable breast cancer by means of statistical modelling based on both retrospective registry data, HES data and prospective cohort study data.
2. To develop a simple scoring system, based on co-morbidity, dependency, age and tumour characteristics, which will enable prediction of those women most likely to benefit from chemotherapy after surgery.
3. To develop a web-based algorithm based on the developed model to aid clinician decision making.

Secondary objectives

1. To determine post-operative surgical outcomes in older women undergoing surgery for breast cancer and correlate outcomes with age, frailty and co-morbidity.
2. To determine chemotherapy adverse events (AEs) in older women undergoing adjuvant chemotherapy for breast cancer and correlate these with patient age, co-morbidity and frailty.
3. To determine quality of life (QoL) outcomes in older women undergoing surgery, chemotherapy or PET for breast cancer and correlate outcomes with age, co-morbidity and frailty.
4. To determine the level of and causes of variance in breast cancer treatment of older women between UK breast units.

Further objectives

1. Data from this study will be made available to collaborators developing a patient Decision Support Instrument (DSI) to facilitate patient decision making by enabling detailed, patient specific outcomes to be predicted.
2. The study has patient consent for long-term access to cancer registry data and outcomes of all women enrolled in the cohort study to permit further longer term analysis of outcomes.

This statistical analysis plan (SAP) is written in conjunction with the International Conference on Harmonisation (ICH) topic E9 [1], applicable standard operating procedures (SOPs) from the University of Sheffield Clinical Trials Research Unit (CTRU) and the trial protocol [2]. The trial will be conducted in accordance with the principles of Good Clinical Practice (GCP) according to the EU Directive 2005/28/EC (GCP Directive), which was implemented in The Medicines for Human Use (Clinical Trials) Amendment Regulations 2006 [3].

## 1.2 Outcome measures

The objectives of the trial will be evaluated on the following endpoints and for PET compared to surgery, and chemotherapy compared to no chemotherapy. All outcome variables will be measured at baseline, 6 weeks, 6, 12, 18 and 24 months post-baseline.

### 1.2.1 Primary outcome

The primary outcome will be a statistical model of outcomes for older women and the determination of the complex and interacting set of characteristics that determine optimal treatment for older women.

### 1.2.2 QOL related secondary outcomes

The following QoL measures are secondary outcomes:

1. The global health status/QoL domain of the EORTC-QLQ-C30 [4], a generic QoL tool, measured at 6 months post-baseline (key secondary outcome)
2. The functional and symptom scales of the EORTC-QLQ-C30, measured at 6 weeks, and then 6 monthly intervals up to 24 months.
3. The EORTC-QLQ-BR23 [5], a breast specific module, measured at 6 weeks, and and then 6 monthly intervals up to 24 months.
4. The EORTC-QLQ-ELD15 [6], an older person specific module, measured at 6 weeks and then 6 monthly intervals up to 24 months.

### 1.2.3 Survival related secondary outcomes

The following outcomes will also be recorded:

1. Cause of and date of death for survival analysis.
2. Disease free survival.
3. Time to local recurrence: time to the relevant event as per the CRF or in the case of death after direct follow up has occurred, as per cancer registry event date recorded.
4. Time to metastatic recurrence: time to the relevant event as per the CRF or in the case of death after direct follow up has occurred, as per cancer registry event date recorded.

### 1.2.4 Other secondary outcomes

1. Date of change of management/treatment. This when disease progression or recurrence occurs mandating the need for a different treatment strategy Change of management (CoM) is when a treatment is abandoned and changed such as when a participant on endocrine therapy is changed to surgery or a different antioestrogen drug. This will be picked on the CRF and the date recorded and the time from the start of that treatment to the date of change is the time to CoM.
2. Time to change of management/treatment: see previous definition.
3. Treatment related AEs (for both chemotherapy, radiotherapy and surgery) and whether these may be predicted by patient variables such as type of surgery, comorbidity, frailty, nutritional state, cognitive function etc.

**Other outputs.**

In addition we will produce a descriptive report of the treatment pathways for women with early breast cancer in this older age group with percentages having the various types of therapies and also look at factors which predict one or other treatment being given or omitted.

The study will also look at variation in rates of surgery or non surgical treatments between units to assess whether this is randomly distributed or whether some centres are significant outliers for normal practice.

## 1.3 Sample size

We propose to recruit and follow-up eligible women from at least 50 UK Breast Units. Each Unit sees between 200 and 700 breast cancers per year, of which 30% will be over age 70. Assuming an uptake rate of 50% in eligible women this will allow us to collect data on over 3500 subjects over the course of the study recruitment period (February 2012 to June 2017). With a median of 2 years of direct follow-up this integrated dataset will provide an evidence base for the medium term post primary treatment. Longer-term follow-up via registry data will maximise the project’s long-term value. We will ask all women to consent for the study team to have access to their registry data and also to give consent for subsequent access to their stored tissue samples (which will form the basis for future research). Details of the sample size for the PET versus surgery and the chemotherapy versus no chemotherapy analysis are given below.

### 1.3.1 PET versus surgery analysis

The study aims to recruit from 50 UK centres. For this analysis women over the age of 75, who might be deemed suitable for either PET or surgery by their clinician, regardless of the treatment they ultimately receive will be eligible if they have ER+ cancers. Women over 75 make up 65% of the recruited population of the study so far, of which 85% will have ER+ cancers. Of the 3000+ recruits 1950 will be over 75 and 1657 will have ER positive disease as well. This group of women may be suitable for a choice of either treatment. National statistics suggest approximately 25% of women have PET if aged over 70 (ranging from 12 to 40% by region). The study will examine the characteristics of this subset of women to determine the characteristics determining type of treatment and outcomes depending on treatment type.

### 1.3.2 Chemotherapy analysis

The standard indications for chemotherapy will be used as a guide to eligibility for the chemotherapy versus no chemotherapy analysis of the study. Women must be over 70 years of age and their cancer must have poor prognostic features (based on the criteria used in the ACTION trial, see below):

Based on data from the Breast Cancer Clinical Outcome Measures (BCCOM) audit 2009, 38% of women between ages 50 and 70 years of age were treated with chemotherapy [7]. Unpublished data from the ACHeW study in women aged 70 or over showed similar rates of women potentially eligible but far fewer actually receiving it: 116 of the 803 patients (14%) were offered chemotherapy as part of their treatment for early breast cancer, with 66 (8%) going on to receive it. Of those 309 women with disease at high risk of recurrence: 94 (30%) were offered chemotherapy, and 53 (17%) received it [A Ring, a personal communication].

On analysis of the database at the 2000 patients recruited point, a total of 175 women have had chemotherapy, the majority under the age of 80, which is very close to the percentage reported in the ACHEW study at 9%. Extrapolating this number to the end of study recruitment when we hope to have recruited between 3000 and 3500 this will give us 270-315 patients actually having chemotherapy and about an equal number who had high risk cancers but did not receive it as a comparator. The study will examine the criteria used to guide selection and the relative cancer and adverse event outcomes in these 2 groups.

## 1.4 Data monitoring

Data will be monitored for quality and completeness by the Study Team. Missing data will be chased until it is received, confirmed as not available, or the trial is at analysis. The study team will conduct source data verification (SDV) on a minimum of 10% of patients.

### 1.4.1 Data Monitoring and Ethics Committee (DMEC)

The DMEC is reviewing data for the study at 6 monthly intervals and is composed of the following independent members: Professor Margot Gosney (a geriatrician), Dr Matthew Hatton (medical oncologist) and Professor Alistair Thompson (breast surgeon).

### 1.4.2 Trial Management Group (TMG)

The TMG meets every 6 months to oversee trial progress, adherence to the protocol, patient safety, and consideration of new information.

### 1.4.3 Interim analysis

There are no statistical criteria for stopping the study early as the study is simply observing normal UK practice and therefore very low risk. At the half way point the study is recruiting at a satisfactory rate and will therefore continue to completion.

Descriptive analyses will be performed and presented to TMG meetings to assess the following:

* Recruitment rates generally and by site and case mix (age subgroup, treatment type)
* data quality and completeness

# 2 Data sources, protocol non-compliance and analysis populations

## 2.1 Data sources

Data used in this study will come from data entered into the following sources:

* Case Report Forms (CRFs)
* Study Questionnaires
* Cancer Registry Outcome Data (longer term outcomes and patients lost to follow up).

The data will be stored on a bespoke database constructed by the Study Data Manager (Mr Tim Chater). Data will be monitored by the study data monitor periodically to check accuracy. SDV will be obtained on 10% of the study sample.

## 2.2 Protocol non-compliances

As this is a cohort study, there are no specific trial related treatment protocols to be followed. We do not expect any protocol non-compliances, as all eligible and consenting women will be included in the analysis provided they have valid outcome data.

## 2.3 Analysis populations

The analysis population will include all recruited women for which consent has been obtained. The study is non-interventional so there are no Intention To Treat (ITT) or Per-Protocol (PP) subpopulations.

# 3 Outline of statistical analyses

## 3.1 General considerations

The EORTC-QLQ-C30, EORTC-QLQ-BR23 and the EORTC-QLQ-ELD15 will be undertaken at baseline and at 6 weeks, 6, 12, 18 and 24 months post-baseline. The EuroQol 5D (EQ-5D) and AEs will also be recorded at baseline and at 6 weeks, 6, 12, 18 and 24 months post-baseline. The following will be recorded at baseline only:

1. Activities of daily living (ADL).
2. Instrumental activities of daily living (IADL).
3. Mini-mental state examination (MMSE).
4. Charlson Index.
5. Eastern Cooperative Oncology Group Performance Status (ECOG PS).

Statistical significance will be taken at the 5% level.

## 3.2 Demographics and baseline characteristics

Baseline socio-demographic (age, ethnicity), tumour characteristics (in the case of bilateral disease the worst tumour will be taken), proportion of subjects experiencing bilateral disease (and listing of tumour characteristics) and individualised baseline scores (EORTC QoL scores, EQ-5D, Barthel, IADL, MMSE, Charlson Index, ECOG PS) will be summarised and assessed for comparability between the different treatment groups (PET versus surgery and chemotherapy versus no chemotherapy). For continuous variables means and standard deviations or medians and interquartile ranges will be calculated depending on the distribution of the data. The number of observations will be presented alongside the summaries. For categorical variables such as age sub-group (75-79, 80-84, 85-89 and 90+ years), comorbidity (based on Charlson Score) and frailty (based on Barthel score) subgroups (and ethnicity but we expect this subgroup analysis will be too small for meaningful analysis in this population), the number and percentage of participants in each of the categories will be presented.

All baseline summaries will be presented and reported for each treatment group (surgery; PET; chemotherapy; no chemotherapy) and in total, as shown in Table 1. Baseline imbalances in these characteristics will be descriptively reported and adjusted for in the statistical model.

### 3.2.1 Definitions and data manipulation

The baseline date is the date of patient consent. The centre will be defined as the place from which the patient was identified.

#### 3.2.1.1 Scoring algorithms of EORTC Instruments

Scoring of the three questionnaires will be integrated into the database. The results will be checked using the raw data by the statistician. For the EORTC-QLQ-C30 and EORTC-QLQ-BR23, the R package QoLR [8] will be used to calculate the scores. This package follows *The EORTC-C30 Scoring Manual (3rd Edition)* [9] for both scoring and dealing with missing data*.* For the EORTC-QLQ-ELD15, R functions will be written in order to calculate the scores.

##### 3.2.1.1.1 EORTC-QLQ-C30

The EORTC-QLQ-C30 and the EORTC-QLQ-BR23 will be scored according to *The EORTC-C30 Scoring Manual (3rd Edition)* [9], which contains information on how to score both questionnaires. A high score for a functional scale represents a high/healthy level of functioning, a high score for the global health status/QoL represents a high QoL, whilst a high score for a symptom scale represents a high level of symptomology/problems. The EORTC-QLQ-C30 comprises 30 items, made up of 9 scales and 6 single items split into three categories: global health status/QoL, functional scales and symptom scales/items. They are as follows:

|  |  |
| --- | --- |
|  | **Items\*** |
| **Global health status/QoL**  Global health status/Qol (revised) | 29, 30 |
| **Functional scales**  Physical functioning (revised)  Role functioning (revised)  Emotional functioning  Cognitive functioning  Social functioning | 1 to 5  6, 7  21 to 24  20, 25  26, 27 |
| **Symptom scales/items**  Fatigue  Nausea and vomiting  Pain  Dyspnoea  Insomnia  Appetite loss  Constipation  Diarrhoea  Financial difficulties | 10, 12, 18  14, 15  9, 19  8  11  13  16  17  28 |

\***Items numbered as in the CRFs**

##### 3.2.1.1.2 EORTC-QLQ-BR23

The EORTC-QLQ-BR23 comprise 23 items, made up of 5 scales and 3 single items split into two categories: functional scales and symptom scales/items. They are as follows:

|  |  |
| --- | --- |
|  | **Items\*** |
| **Functional scales**  Body image  Sexual functioning  Sexual enjoyment  Future perspective | 39 to 42  44, 45  46  43 |
| **Symptom scales/items**  Systemic therapy side effects  Breast symptoms  Arm symptoms  Upset by hair loss | 31 to 34, 36, 37, 38  50 to 53  47, 48, 49  35 |

**\*Items numbered as in the CRFs**

The scoring manual [9] also discusses the method for dealing with missing data, which is described in Section 3.5.

##### 3.2.1.1.2 EORTC-QLQ-ELD14

The EORTC-QLQ-ELD15 uses the same scoring system as for the EORTC-QLQ-C30 and the EORTC-QLQ-BR23 [9]. It comprises 15 items, made up of 5 scales as shown below:

|  |  |
| --- | --- |
|  | **Items\*** |
| Mobility  Family support  Worries about the future  Maintaining autonomy and purpose  Burden of illness | 31 to 34  35, 36  37 to 41  42, 43  44, 45 |

**\*Items numbered as in the CRFs**

However, following an international validation study [10], one item (Question 35 - Has your relationship with your family become closer?) was removed from the EORTC-QLQ-ELD15, resulting in the EORTC-QLQ-ELD14, and the scale structure was revised. The EORTC-QLQ-ELD14 comprises 14 items, made up of 5 scales and 2 single items as shown below:

|  |  |
| --- | --- |
|  | **Item\*** |
| Mobility  Joint stiffness  Family support  Worries about others  Future worries  Maintaining purpose  Burden of illness | 31, 33, 34  32  36  37, 38  39 to 41  42, 43  44, 45 |

**\*Items numbered as for the ELD15 in the CRFs**

We will use the new EORTC-QLQ-ELD14 scoring system to score the EORTC-QLQ-ELD15 (i.e. question 35 will be removed and the new scale structure followed). The responses will be scored in the same way as for the EORTC-QLQ-C30 and the EORTC-QLQ-BR23 [9], to report QoL on a scale of 0-100 where 100 indicates better QoL [11]. Therefore, mobility, joint stiffness, worries about others, future worries and burden of illness will be scored using the functional scale as defined in [9], whilst family support and maintaining purpose will be scored using the symptom scale as defined in [9].

#### 3.2.1.2 EQ-5D

PLACEHOLDER: this will mainly be used for the health economics stuff

#### 3.2.1.3 Barthel ADL

Instructions: Choose the scoring point for the statement that most closely corresponds to the patient's current level of ability for each of the following 10 items. Record actual, not potential, functioning. Information can be obtained from the patient's self-report, from a separate party who is familiar with the patient's abilities (such as a relative), or from observation. Refer to the Guidelines section on the following page for detailed information on scoring and interpretation.

The Barthel Index

Bowels

0 = incontinent (or needs to be given enemata)

1 = occasional accident (once/week)

2 = continent

Patient's Score:

Bladder

0 = incontinent, or catheterized and unable to manage

1 = occasional accident (max. once per 24 hours)

2 = continent (for over 7 days)

Patient's Score:

Grooming

0 = needs help with personal care

1 = independent face/hair/teeth/shaving (implements provided)

Patient's Score:

Toilet use

0 = dependent

1 = needs some help, but can do something alone

2 = independent (on and off, dressing, wiping)

Patient's Score:

Feeding

0 = unable

1 = needs help cutting, spreading butter, etc.

2 = independent (food provided within reach)

Patient's Score:

Transfer

0 = unable – no sitting balance

1 = major help (one or two people, physical), can sit

2 = minor help (verbal or physical)

3 = independent

Patient's Score:

Mobility

0 = immobile

1 = wheelchair independent, including corners, etc.

2 = walks with help of one person (verbal or physical)

3 = independent (but may use any aid, e.g., stick)

Patient's Score:

Dressing

0 = dependent

1 = needs help, but can do about half unaided

2 = independent (including buttons, zips, laces, etc.)

Patient's Score:

Stairs

0 = unable

1 = needs help (verbal, physical, carrying aid)

2 = independent up and down

Patient's Score:

Bathing

0 = dependent

1 = independent (or in shower)

Patient's Score:

Total Score:

(Collin et al., 1988)

Scoring:

Sum the patient's scores for each item. Total possible scores range from 0 – 20, with lower scores indicating increased disability. If used to measure improvement after rehabilitation, changes of more than two points in the total score reflect a probable genuine change, and change on one item from fully dependent to independent is also likely to be reliable.

For subgroup analysis Barthel ADL will be categorised to: mild, moderate, severe. Find below the definitions:

Mild:

Moderate:

Severe:

Sources:

• Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: a reliability study. Int Disabil Stud. 1988;10(2):61-63.

• Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J. 1965;14:61-65.

• Wade DT, Collin C. The Barthel ADL Index: a standard measure of physical disability? Int Disabil Stud. 1988;10(2):64-67.

#### 3.2.1.4 IADL

#### The patient receives a score of 1 for each item labeled A – H if his or her competence is rated at some minimal level or higher. Add the total points circled for A – H. The total score may range from 0 – 8. A lower score indicates a higher level of dependence. In case of missing items the overall score is set to missing3.2.1.4 MMSE

The following three cut-off levels should be employed to classify the severity of cognitive impairment: no cognitive impairment=24-30; mild cognitive impairment=18-23; severe cognitive impairment=0-17. In case of missing data …..

#### 3.2.1.6 Charlson score

Each condition is assigned a score of 1, 2, 3, or 6, depending on the risk of dying associated with each one. Scores are summed to provide a total score to predict mortality. Many variations of the Charlson comorbidity index have been presented, including the Charlson/Deyo, Charlson/Romano, Charlson/Manitoba, and Charlson/D'Hoores comorbidity indices.

Clinical conditions and associated scores are as follows:

* 1 each: Myocardial infarct, congestive heart failure, peripheral vascular disease, dementia, cerebrovascular disease, chronic lung disease, connective tissue disease, ulcer, chronic liver disease, diabetes.
* 2 each: Hemiplegia, moderate or severe kidney disease, diabetes with end organ damage, tumor, leukemia, lymphoma.
* 3 each: Moderate or severe liver disease.
* 6 each: Malignant tumor, metastasis, AIDS.

The next step was to stratify the risk scores into four risk groups: (1) score of 0 to 3; (2) score of 4 to 7; (3) score of 8 to 11 and (4) score of 12 and above. From https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4744039/

Find below the categories and definition for subgroup analysis:

Category 1:

Category 2:

#### 3.2.1.7 ECOG Performance Status

Grade 0: Fully active, able to carry on all pre-disease performance without restriction

Grade 1: Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work

Grade 2: Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours

Grade 3: Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours

Grade 4: Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair

Grade 5: Dead

### 3.2.2 Recruitment and data completeness

A CONSORT style flow diagram will be used to display data completeness and patient throughput from eligibility screening, invitation, study acceptance and final follow-up visit. The following summary statistics will also be reported:

1. The number of patients screened per month.
2. The number of patients recruited per month.
3. The number and percentage of patients who complete each follow up visit or who are lost to follow up.
4. The number of patients who have complete data for each key variable.

To allow time for data entry, items will only be considered incomplete if they have not been entered within 30 days of the expected date.

## 3.3 Clinical outcomes

### 3.3.1 QOL key secondary outcome

The QOL key secondary outcome will be the global health status/QoL domain of the EORTC-QLQ-C30, measured at 6 months.

### 3.3.2 Statistical analysis for the QOL key secondary outcome

Since the study is a cohort it is likely that the baseline demographic, clinical and QoL characteristics of the women on the different treatment regimens (surgery versus PET, and chemotherapy versus no chemotherapy) are different and this may influence future outcomes. In order to make sure that we are comparing like with like and allow for differences in case-mix between the different treatment regimens we shall use a variety of statistical methods. The two main statistical approaches that will be used to adjust for baseline imbalances in patient characteristics will be:

1. Analysis of covariance (ANCOVA) and
2. Propensity score methods.

With sufficiently large numbers of patients an ANCOVA model alone is often sufficient and so this will be the primary analysis method. However, ANCOVA can produce biased estimates of treatment effect if there is extreme imbalance in baseline characteristics or if the treatment effect is not constant with respect to the baseline characteristics. We will therefore also use propensity score methods as a sensitivity analysis. The propensity score methods are based on determining an individual patient’s probability of being treated, with a particular therapy/regimen, conditional on their baseline characteristics. The two methods are described in detail below.

ANCOVA

Separate ANCOVA models will be fitted for the PET versus surgery and chemotherapy versus no chemotherapy groups for the QOL key secondary outcome (global health status/QoL domain of the EORTC-QLQ-C30 at 6 months). The models will be adjusted for baseline global health status EORTC-QLQ-C30 score and will include treatment group and any other clinically important baseline covariates.

Propensity score methods

A propensity score for each patient or the probability of having a particular treatment regimen (e.g. PET or surgery) will be derived from a binary logistic regression model. A second propensity score for each patient or the probability of being treated with chemotherapy or no chemotherapy will also be derived from a binary logistic regression model. For both models the outcome will be treatment group and the covariates to be included will be baseline variables that are either potential confounders or predictors of treatment group. The covariates will be chosen based on expert knowledge and will relate to both treatment and outcome.

We will then use matching, stratification and ANCOVA using the calculated propensity scores (for each individual patient) to balance patient characteristics between treatments and allow the estimation of an unbiased estimate of treatment effect of firstly PET versus surgery and secondly chemotherapy versus no chemotherapy on the primary outcome. The process of matching, stratification and ANCOVA are discussed below.

*Matching*

We will use greedy nearest neighbour caliper matching without replacement, as suggested in Austin [12]. The process is as follows and can be performed in a number of statistical packages:

1. Select a treated subject at random (i.e. a patient that receives PET in the PET versus surgery analysis or a patient that receives chemotherapy in the chemotherapy versus no chemotherapy analysis).
2. Identify all untreated subjects whose propensity score is within a given caliper distance of the chosen subject’s propensity score. The caliper distance used will be 0.2 of the standard deviation of the logit of the (treated subject’s) propensity score, as suggested in the literature [13]. From this restricted set of untreated subjects, choose the untreated subject with the closest propensity score to the treated subject as a match. If multiple untreated subjects have a propensity score within equal distance of the treated subject, select the untreated subject at random.
3. If no untreated subjects lie within the caliper distance then no match is made and the treated subject is excluded from the matched sample.
4. Once an untreated subject has been matched to a treated subject they are no longer available as a potential match for subsequent treated subjects (matching without replacement).

Steps 1 to 5 are repeated until untreated subjects are matched to all treated subjects or there are no treated subjects remaining for whom a matched untreated subject can be found [14].

The average effect of treatment in those subjects who ultimately receive the treatment (ATT) can then be estimated by calculating the difference in mean global health status/QoL EORTC-QLQ-C30 scores between the treated and untreated matches.

*Stratification*

The process of stratification is as follows and can be performed in a number of statistical packages:

1. Rank the subjects according to their propensity score.
2. Stratify the subjects into subsets. The subsets will be based on the quintiles of the estimated propensity scores [14]. Within each stratum treated and untreated subjects will have similar propensity scores.

We then estimate the effect of treatment within each stratum by calculating the difference in mean global health status/QoL EORTC-QLQ-C30 scores between treated and untreated subjects. To estimate the overall treatment effect the stratum specific effects are pooled. For estimation of the ATT, weight the stratum-specific estimates by the proportion of treated subjects within each stratum.

*ANCOVA/covariate adjustment*

Separate ANCOVA models for the PET versus surgery and chemotherapy versus no chemotherapy groups will be fitted for the primary outcome (global health status/Qol EORTC-QLQ-C30 score at 6 months). The model will include only treatment group and the calculated propensity score for each individual and will be of the form:

where =primary outcome (global health status/QoL EORTC-QLQ-C30 score), treatment group and is the propensity score. The estimate of is the estimate of the exposure/treatment effect.

### 3.3.3 Quality of life analysis

An initial descriptive and graphical analysis will summarise mean global health status/QoL score of the EORTC-QLQ-C30 questionnaire over time (baseline, 6, 12, 18 and 24 months follow-up) by treatment group to examine the pattern of change over time. Mean global health status/QoL score and standard deviation for the EORTC-QLQ-C30 at each time point (baseline, 6, 12, 18 and 24 months follow-up) will be presented as shown in Table 2. Spaghetti plots will be produced separately for the PET versus surgery and chemotherapy versus no chemotherapy groups. Results will be presented for complete cases (global health status/QoL EORTC-QLQ-C30 score available at all time points) as well as all available cases at each time point (varying total number of patients at each time point).

A series of longitudinal models will be used to compare repeated global health status/QoL EORTC-QLQ-C30 scores over time (baseline, 6, 12, 18 and 24 months follow-up) between women treated with different therapies (PET versus surgery; or chemotherapy versus no chemotherapy). The longitudinal model will be a random effect (multi-level) general linear model [13]. The longitudinal model will allow for time, treatment, treatment-time interaction, and adjust for baseline global health status/QoL EORTC-QLQ-C30 scores and other baseline prognostic covariates or propensity score.

Mean global health status/QoL EORTC-QLQ-C30 scores and 95% confidence intervals (CIs) adjusted for baseline score or propensity score will be calculated for all domains at each time point, as shown in Table 4 and Table 5. The EORTC-QLQ-C30 forms will be scrutinised for missing data, and missing data patterns investigated. Section 3.5 discusses the handling of missing data in more detail.

The mean scores (standard deviations) of each domain of the EORTC-QLQ-C30, EORTC-QLQ-BR23 and the EORTC-QLQ-ELD15 at each time point will be presented as shown in Table 2/Table 3, Table 6/Table 7 and Table 8/Table 9, respectively.

The mean scores and 95% CIs of each domain of the EORTC-QLQ-C30, EORTC-QLQ-BR23 and the EORTC-QLQ-ELD15 adjusted for baseline score or propensity score will be presented as shown in Table 4/Table 5, Table 10/Table 11 and Table 12/Table 13, respectively.

### 3.3.4 Analysis of survival related outcomes

The following summary statistics will be reported overall and by treatment type (PET, surgery, chemotherapy, no chemotherapy):

1. Number of deaths.
2. Number of deaths by cause (breast specific or other).
3. (for other time to event outcomes) Number of events under consideration

Kaplan-Meier curves will be derived for each treatment type, by age, disease characteristics, co-morbidity and frailty subgroups (to be defined) to illustrate how these factors interact with treatment type and disease stage.

OS will primarily be compared between the treatment groups using multivariate modelling, Cox’s Proportional Hazards model if appropriate. OS will also be compared using multivariate modelling adjusting for important prognostic factors. Hazard ratios (HR) and corresponding 95% CIs will be presented. Yearly OS, median survival, and corresponding 95% CIs will be presented for each treatment type.

Addititionally we’ll report for the model based on OS [19,20]

- Calibration and Discrimination Indexes/graphs. For Calibration we’ll produce graphs of predicted survival probabilities against right-censored failure times [21]. For discrimination, for example [19], we can produce the following statistics: Harrell c-index, Gonen & Heller K, Explained variation (R^2\_D).

- the untransformed coefficients [19]

- Kaplan Meier estimates grouped by Prognostic Index [19]

- the baseline survival curve and baseline survival for useful timepoints [19]

Failure-free survival (FFS) and cumulative incidence functions for local disease control will be calculated and compared using multivariate modelling, Cox’s Proportional Hazards model if appropriate to adjust for any important prognostic factors. HRs and corresponding 95% CIs will be presented. Yearly survival/local disease control, median survival/local disease control, and corresponding 95% CIs will be presented for each treatment group. Time to local progression or recurrence and its treatment will be summarised descriptively.

### 3.3.5 Subgroup evaluation

Age subgroup analysis (75-79, 80-84, 85-89, 90+), Barthel Index subgroup analysis (mild, moderate, severe), Charlson score and degree of dementia (mild, moderate or severe) will also be carried out on the QOL and survival secondary outcome measures.

A plot of mean score for each domain of the QoL outcomes (EORTC-QLQ-C30, BR23, ELD15) at 6 months against subgroup will be produced. A subgroup by treatment interaction term will be fitted to the ANCOVA, longitudinal (random effects) and survival models (Cox Proportional Hazards) - see Sections 3.3.1.1, 3.3.1.2, 3.3.1.3 and 3.3.3 - to test for a significant difference in outcome between subgroups. That is to say, the models will include treatment group and subgroup, and the interaction term between treatment and subgroup will then be added to test for a significant difference. We will test for subgroup effects even if the main treatment effect is insignificant.

### 3.3.6 Model diagnostics

Model diagnostics will be performed only on the QOL key secondary outcome.

ANCOVA

Homogeneity of variance will be assessed by plotting the studentised residuals against the predicted values from the model, whilst Normality will be assessed using Normal probability plots. If the assumptions for the ANCOVA are violated either an appropriate transformation will be applied or a non-parametric procedure with less stringent assumptions will be utilised as sensitivity analysis.

Propensity score model

The fit of the propensity score model is checked by assessing the balance in covariates between treatment groups. The diagnostic checks for each propensity score method are described below.

*Matching*

A comparison of means/medians of the continuous covariates and distribution of the categorical variables used in the propensity score model will be made between treated and untreated subjects.

Standardized differences will be calculated to compare the mean of continuous and binary variables between treatment groups. For continuous variables the standardized difference is given by

where and denote the sample mean of the covariate in the treated and untreated subjects, respectively, and and denote the sample variance of the covariate in the treated and untreated subjects, respectively [12]. For the categorical variables, the standardized difference is given by

where and denote the prevalence or mean of the categorical variable in treated and untreated subjects, respectively [10]. A standardized difference of less than 0.1 will be taken to indicate a negligible difference in the mean or prevalence of a covariate between groups [10]. If the standardized difference is greater than 0.1, the propensity score model will be revised.

Graphical methods (e.g. boxplots, quantile-quantile plots) will also be utilised to assess the balance in covariates between the treatment groups.

If important systematic differences are found, the original propensity score model will be adjusted by either: including additional covariates, adding interaction terms between existing covariates, or by modelling the relationship between continuous covariates and treatment group using non-linear terms. An iterative process will be performed (calculating propensity scores, matching, model checks) until systematic differences in covariates have been removed or reduced to an acceptable level.

*Stratification*

The diagnostic checks used for the matching method will also be utilised for the stratification method; however, the checks are performed for each stratum in this case.

Again, if important systematic differences are found, the original propensity score model will be adjusted by either: including additional covariates, adding interaction terms between existing covariates, or by modelling the relationship between continuous covariates and treatment status using non-linear terms. The number of strata may also be revised in order to balance covariates between treatment groups within each stratum. An iterative process will be performed (calculating propensity scores, stratification, model checks) until systematic differences in covariates have been removed or reduced to an acceptable level.

*ANCOVA/covariate adjustment*

Weighted conditional standardized differences (an extension of standardized differences) will be calculated to compare the difference in means of baseline covariates between treated and untreated subjects [16]. Quantile regression models will be used to qualitatively compare the distribution of measured baseline covariates between treated and untreated subjects with the same propensity score [16].

## 3.4 Safety outcomes

Adverse events (AEs) will be reported as number and percentage of patients overall and compared between treatment groups but no formal statistical analysis is planned. Treatment related adverse events and reasons for stopped treatment will be summarised descriptively for each treatment group. The maximum grade of toxicity (AEs) per patient and the overall rate of toxicities will be summarised according to standard Common Terminology Criteria for Adverse Events (CTCAE). Analysis of risk factors for adverse events, such as age subgroup, comorbidity or frailty subgroup will also be performed. ? analysis by major or minor surgery as well as per Osama’s project and systemic or local complications

## 3.5 Missing and spurious data

Any spurious data will be queried and checked for consistency with data management before data lock. We do not intend to impute any missing data for the cohort study other than that required for calculation of the propensity scores. If baseline covariates required for calculating the propensity scores are missing, we will impute values using multiple imputation.

Missing items

Details for dealing with missing items in the three QoL questionnaires (EORTC-QLQ-C30, BR23 and ELD15) are documented in *The EORTC QLQ-C30 Scoring Manual (3rd Edition)* [9]. In summary, the following imputation method is used:

* Have at least half of the items from the scale been answered?
* If *Yes*, use all the items that were completed and apply the standard equations for calculating the scale scores; ignore any items with missing values when making the calculations.
* If *No*, set scale score to missing.
* For single item measures, set score to missing.

## 3.6 Health economic analyses

A health economic analysis is planned to compare the cost-effectiveness of the following treatment strategies;

* For women with operable ER+ disease; surgery (with adjuvant endocrine therapy) versus primary endocrine therapy
* For women treated surgically with high risk of recurrence; surgery with adjuvant chemotherapy versus surgery without chemotherapy

This work will be led by the team based in Health Economics and Decision Sciences within the School of Health and Related Research at the University of Sheffield. A conceptual framework for both treatment decision problems will be developed on the basis of evidence from the medical literature in consultation with medical experts and members of the Bridging the Age Gap research team. This model will adopt a lifetime horizon from the time of breast cancer diagnosis, and will incorporate factors influencing the treatment decision, health outcomes, patient preferences and costs.

The conceptual model will be used to implement a mathematical model to compare outcomes and costs for the two treatment decision problems. The primary analysis will be a cost-utility analysis (CUA), which will estimate the health outcomes (measured in Quality of Life Years (QALYs)) achieved and the health service costs incurred for patients receiving different treatment options. The model will be an individual level patient simulation, with outcomes dependent on the individual characteristics of the patient. Estimates of short term risk of disease relapse, adverse events and survival, as well as health utility and resource use will be derived from the statistical analyses of the observational study described elsewhere in this document. Long term outcomes will be extrapolated by considering other evidence, which will include retrospective cancer registration data and evidence from the literature.

This analysis will be used to estimate the Incremental Cost Effectiveness Ratio (ICER) for surgery vs PET and for adjuvant vs no chemotherapy for different patient subgroups defined in terms of their age, underlying health status and disease characteristics. Cost-effectiveness will be assessed by comparing the estimated ICERs with the willingness-to-pay thresholds used by the National Institute for Health and Care Excellence (NICE). Uncertainty in model outputs will be evaluated using probabilistic sensitivity analysis (PSA), with uncertainty in model inputs incorporated using probability distributions which will be derived from the statistical analysis of the observational study and other evidence where appropriate.

# 4 Appendix

## 4.1 Tables Output

Table 1 Demographic and baseline characteristics of patients at baseline.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **PET**  **(n=XX)** | **Surgery**  **(n=XX)** | **Chemotherapy**  **(n=XX)** | **No chemotherapy**  **(n=XX)** | **All**  **(n=XX)** |
| Age |  |  |  |  |  |
| Age group  70-74  75-79  80-84  85-89  90+ |  |  |  |  |  |
| Ethnicity  White  Black  Mixed  Asian  Other |  |  |  |  |  |
| EORTC-QLQ-C30  Global health status/QoL  Physical functioning  Role functioning  Emotional functioning  Cognitive functioning  Social functioning  Fatigue  Nausea and vomiting  Pain  Dyspnoea  Insomnia  Appetite loss  Constipation  Diarrhoea  Financial difficulties |  |  |  |  |  |
| EORTC-QLQ-BR23  Body image  Sexual functioning  Sexual enjoyment  Future perspective  Systemic therapy side effects  Breast symptoms  Arm symptoms  Upset by hair loss |  |  |  |  |  |
| EORTC-QLQ-ELD14 \*  Mobility  Joint stiffness  Family support  Worries about others  Maintaining purpose  Burden of illness |  |  |  |  |  |
| EQ-5D |  |  |  |  |  |
| ADL |  |  |  |  |  |
| IADL |  |  |  |  |  |
| MMSE |  |  |  |  |  |
| Charlson Index |  |  |  |  |  |
| ECOG PS |  |  |  |  |  |
| Tumour characteristics:  N only left  N only right  N both left and right |  |  |  |  |  |
| Tumour characteristics (if bilateral, summarise worst)  Stage  Grade  nodel status  tumour size  ER  Her 2 receptor status  surgery or not  type of surgery  other treatment percentages  chemotherapy or not |  |  |  |  |  |

\*Formerly EORTC-QLQ-ELD15.

Table 2 Mean scores (standard deviations) for each domain of the EORTC-QLQ-C30 at each time point: PET versus surgery\*.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Time point**  **(months)** | **Domain** | **PET** | **Surgery** | **n** | **Effects** |
| 0 (baseline)  6  12  18  24 | Global health status/QoL | XX (XX)  … | XX (XX)  … | XX  … | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Physical functioning |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Role functioning |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Emotional functioning |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Cognitive functioning |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Social functioning |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Fatigue |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Nausea and vomiting |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Pain |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| **Time point**  **(months)** | **Domain** | **PET** | **Surgery** | **n** |  |
| 0 (baseline)  6  12  18  24 | Dyspnoea |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Insomnia |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Appetite loss |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Constipation |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Diarrhoea |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Financial difficulties |  |  |  | N=  Treatment=  Time=  Treatment\*time= |

\*The table will be repeated for both complete cases (non-varying n) and all available cases (varying n at each time point).

Table 3 Mean scores (standard deviation) for each domain of the EORTC-QLQ-C30 at each time point: chemotherapy versus no chemotherapy\*.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Time point**  **(months)** | **Domain** | **Chemotherapy** | **No chemotherapy** | **n** | **Effects** |
| 0 (baseline)  6  12  18  24 | Global health status/QoL | XX (XX)  … | XX (XX)  … | XX  … | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Physical functioning |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Role functioning |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Emotional functioning |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Cognitive functioning |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Social functioning |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Fatigue |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Nausea and vomiting |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Pain |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| **Time point**  **(months)** | **Domain** | **Chemotherapy** | **No Chemotherapy** | **n** |  |
| 0 (baseline)  6  12  18  24 | Dyspnoea |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Insomnia |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Appetite loss |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Constipation |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Diarrhoea |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Financial difficulties |  |  |  | N=  Treatment=  Time=  Treatment\*time= |

\*The table will be repeated for both complete cases (non-varying n) and all available cases (varying n at each time point).

Table 4 Mean scores and 95% confidence intervals (CIs) adjusted for baseline score or propensity scores for the EORTC-QLQ-C30 at 6 weeks\*: PET versus surgery.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Domain** | **PET**  **(n=XX)** | **Surgery**  **(n=XX)** | **Treatment effect** | **P value** |
| **C30** | Global health status/QoL  Physical functioning  Role functioning  Emotional functioning  Cognitive functioning  Social functioning  Fatigue  Nausea and vomiting  Pain  Dyspnoea  Insomnia  Appetite loss  Constipation  Diarrhoea  Financial difficulties | XX (XX, XX)  … | XX (XX, XX)  … |  |  |

\* The table will be repeated for the 6, 12, 18 and 24 month time points.

Table 5 Mean scores and 95% confidence intervals (CIs) adjusted for baseline score or propensity score for the EORTC-QLQ-C30 at 6 weeks\*: chemotherapy versus no chemotherapy.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Domain** | **Chemotherapy**  **(n=XX)** | **No Chemotherapy**  **(n=XX)** | **Treatment effect** | **P value** |
| **C30** | Global health status/QoL  Physical functioning  Role functioning  Emotional functioning  Cognitive functioning  Social functioning  Fatigue  Nausea and vomiting  Pain  Dyspnoea  Insomnia  Appetite loss  Constipation  Diarrhoea  Financial difficulties | XX (XX, XX)  … | XX (XX, XX)  … |  |  |

\* The table will be repeated for the 6, 12, 18 and 24 month time points.

Table 6 Mean scores (standard deviation) for each domain of the EORTC-QLQ-BR23 at each time point: PET versus surgery\*.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Time point**  **(months)** | **Domain** | **PET** | **Surgery** | **n** | **Effects** |
| 0 (baseline)  6  12  18  24 | Body image | XX (XX)  … | XX (XX)  … | XX  … | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Sexual functioning |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Sexual enjoyment |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Future perspective |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Systemic therapy side effects |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Breast symptoms |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Arm symptoms |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Upset by hair loss |  |  |  | N=  Treatment=  Time=  Treatment\*time= |

\*The table will be repeated for both complete cases (non-varying n) and all available cases (varying n at each time point).

Table 7 Mean scores (standard deviation) for each domain of the EORTC-QLQ-BR23 at each time point: chemotherapy versus no chemotherapy\*.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Time point**  **(months)** | **Domain** | **Chemotherapy** | **No chemotherapy** | **n** | **Effects** |
| 0 (baseline)  6  12  18  24 | Body image | XX (XX)  … | XX (XX)  … | XX  … | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Sexual functioning |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Sexual enjoyment |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Future perspective |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Systemic therapy side effects |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Breast symptoms |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Arm symptoms |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Upset by hair loss |  |  |  | N=  Treatment=  Time=  Treatment\*time= |

\*The table will be repeated for both complete cases (non-varying n) and all available cases (varying n at each time point).

Table 8 Mean scores (standard deviation) for each domain of the EORTC-QLQ-ELD14 (formerly ELD15) at each time point: PET versus surgery\*.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Time point**  **(months)** | **Domain** | **PET** | **Surgery** | **n** | **Effects** |
| 0 (baseline)  6  12  18  24 | Mobility | XX (XX)  … | XX (XX)  … | XX  … | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Joint stiffness |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Family support |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Worries about others |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Maintaining purpose |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Burden of illness |  |  |  | N=  Treatment=  Time=  Treatment\*time= |

\*The table will be repeated for both complete cases (non-varying n) and all available cases (varying n at each time point).

Table 9 Mean scores (standard deviation) for each domain of the EORTC-QLQ-ELD14 (formerly ELD15) at each time point: chemotherapy versus no chemotherapy\*.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Time point**  **(months)** | **Domain** | **Chemotherapy** | **No chemotherapy** | **n** | **Effects** |
| 0 (baseline)  6  12  18  24 | Mobility | XX (XX)  … | XX (XX)  … | XX  … | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Joint stiffness |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Family support |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Worries about others |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Maintaining purpose |  |  |  | N=  Treatment=  Time=  Treatment\*time= |
| 0 (baseline)  6  12  18  24 | Burden of illness |  |  |  | N=  Treatment=  Time=  Treatment\*time= |

\*The table will be repeated for both complete cases (non-varying n) and all available cases (varying n at each time point).

Table 10 Mean scores and 95% confidence intervals (CIs) adjusted for baseline score or propensity score for the EORTC-QLQ-BR23 at 6 weeks\*: PET versus surgery.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Domain** | **PET**  **(n=XX)** | **Surgery**  **(n=XX)** | **Treatment effect** | **P value** |
| Body image  Sexual functioning  Sexual enjoyment  Future perspective  Systemic therapy side effects  Breast symptoms  Arm symptoms  Upset by hair loss | XX (XX, XX)  … | XX (XX, XX)  … |  |  |

\* The table will be repeated for the 6, 12, 18 and 24 month time points.

Table 11 Mean scores and 95% confidence intervals (CIs) adjusted for baseline score or propensity score for the EORTC-QLQ-BR23 at 6 weeks\*: chemotherapy versus no chemotherapy.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Domain** | **Chemotherapy**  **(n=XX)** | **No chemotherapy**  **(n=XX)** | **Treatment effect** | **P value** |
| Body image  Sexual functioning  Sexual enjoyment  Future perspective  Systemic therapy side effects  Breast symptoms  Arm symptoms  Upset by hair loss | XX (XX, XX)  … | XX (XX, XX)  … |  |  |

\* The table will be repeated for the 6, 12, 18 and 25 month time points.

Table 12 Mean scores and 95% confidence intervals (CIs) adjusted for baseline score or propensity score for the EORTC-QLQ-ELD14 (formerly ELD15) at 6 weeks\*: PET versus surgery.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Domain** | **PET**  **(n=XX)** | **Surgery**  **(n=XX)** | **Treatment effect** | **P value** |
| Mobility  Joint stiffness  Family support  Worries about others  Maintaining purpose  Burden of illness | XX (XX, XX)  … | XX (XX, XX)  … |  |  |

\* The table will be repeated for the 6, 12, 18 and 24 month time points.

Table 13 Mean scores and 95% confidence intervals (CIs) adjusted for baseline score or propensity score for the EORTC-QLQ-ELD14 (formerly ELD15) at 6 weeks\*: chemotherapy versus no chemotherapy.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Domain** | **Chemotherapy**  **(n=XX)** | **No chemotherapy**  **(n=XX)** | **Treatment effect** | **P value** |
| Mobility  Joint stiffness  Family support  Worries about others  Maintaining purpose  Burden of illness | XX (XX, XX)  … | XX (XX, XX)  … |  |  |

\* The table will be repeated for the 6, 12, 18 and 24 month time points.

# 5 References

[1] ICH Harmonised Tripartite Guideline Topic E9: Statistical Principles for Clinical Trials 1998. Available at:

<http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf>

[2] Bridging the Age GAP Protocol Version 3, 31 July 2013

[3] The Medicines for Human Use (Clinical Trials) Amendment (No.2) Regulations 2006

[4] Aaronson, N.K., et al. The European Organisation for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute:* 1993; 85: 363-376

[5] Sprangers, M.A., Groenvold, M., Arraras, J.I., Franklin, J., te Velde, A., Muller, M., et al. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. *Journal of Clinical Oncology*: 1996; 14(10): 2756-2768.

[6] Johnson, C., Fitzsimmons, D., Gilbert, J., Arraras, J.I., Hammerlid, E., Bredart, A., et al. Development of the European Organisation for Research and Treatment of Cancer quality of life questionnaire module for older people with cancer: The EORTC QLQ-ELD15. *European Journal of Cancer*: 46(12): 2010; 2242-2252.

[7] Surgery AoBSatBAfC. Breast cancer clinical outcome measures, (BCCOM) 2007; 26. Available at:

<http://www.wmciu.nhs.uk/documents/BCCOM%20Year%203%20report.pdf>

[8] Anota, A. Analysis of Health-Related Quality of Life in oncology. *R package version 1.0.* 2014.

[9] Fayers, P. M., Aaronson, N. K., Bjordal, K., Groenvold, M., Curran, D., and Bottomley, A. The EORTC QLQ-C30 Scoring Manual (3rd Edition). Brussels, *European Organisation for Research and Treatment of Cancer.* 2001.

[10] Wheelwright, S., et al. International validation of the EORTC QLQ-ELD14 questionnaire for assessment of health-related quality of life elderly patients with cancer. *British Journal of Cancer:* 2013; 109: 852-58.

[11] Johnson, C., et al. Development of the European Organisation for Research and Treatment of Cancer quality of life questionnaire module for older people with cancer: The EORTC QLQ-ELD15. *European Journal of Cancer:* 2010; 46: 2242-2252.

[12] Austin, P. C. A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine:* 2013; 33(6): 1057–1069.

[13] Austin, P.C. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions of observational studies. *Pharmaceutical Statistics:* 2011; 10: 150-161.

[14] Austin, P.C. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioural Research:* 2011; 46: 399-424.

[15] Walters, S.J. Quality of Life Outcomes in Clinical Trials and Health-Care Evaluation: a Practical Guide to Analysis and Interpretation. *John Wiley & Sons.* 2009.

[16] Austin, P.C., Goodness-of-fit diagnostics for the propensity score model when estimating treatment effects using covariate adjustment with the propensity score. *Pharmacoepidemiology and Drug Safety:* 2008; 17: 1202-1217.

[17] Thernau, T.M. and Grambsch, P.M. Modeling Survival data analysis: Extending the Cox Model. *Springer Science & Business Media.* 2000.

[18] Bradburn, M.J., et al. Survival analysis Part III: multivariate data analysis - choosing a model and assessing its adequacy and fit. *British Journal of Cancer:* 2003; 89: 605.

[19] Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol. 2013 Mar 6;13:33. doi: 10.1186/1471-2288-13-33. PubMed PMID: 23496923; PubMed Central PMCID: PMC3667097.

[20] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015 Jan 6;162(1):W1-73. doi: 10.7326/M14-0698. PubMed PMID: 25560730.

[21] RMS package <https://cran.r-project.org/web/packages/rms/>