  

**Statistical Analysis Plan**

**Draft Version 0.4**

**02nd February 2017**

Study Title **Bridging the Age GAP in Breast Cancer: Improving outcomes for older women. Evaluation of a decision support intervention for older women with operable breast cancer. A cluster RCT nested within the Age Gap 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*

*CTRU, University of Sheffield*

Approved by

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

*Stephen Walters*

*Professor of Medical Statistics and Clinical Trials*

*ScHARR, University of Sheffield*

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

*Lynda Wyld*

*Reader in Surgical Oncology, Chief Investigator and TMG Chair*

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

*…..*

*TSC chair*

SAP HISTORY

|  |  |  |  |
| --- | --- | --- | --- |
| **Version Date** | **Date**  **approved** | **Modifications (with section)** | **Prior to/After blind review**  **Prior to/after database lock** |
| 0.4 |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Table of Contents**

[1 Introduction, study design and key trial objectives 5](#_Toc428171769)

[1.1 Study outline 5](#_Toc428171770)

[1.2 Hypotheses 6](#_Toc428171771)

[1.3 Outcome measures 6](#_Toc428171772)

[1.3.1 Primary endpoints 6](#_Toc428171773)

[1.3.2 Key secondary endpoints 6](#_Toc428171774)

[1.3.2 Secondary endpoints 6](#_Toc428171775)

[1.3.3 Other outcomes 7](#_Toc428171776)

[1.3.4 Process evaluation measures 7](#_Toc428171777)

[1.4 Sample size 7](#_Toc428171778)

[1.5 Randomisation 7](#_Toc428171779)

[1.6 Data monitoring 8](#_Toc428171780)

[1.6.1 Data Monitoring and Ethics Committee (DMEC) 8](#_Toc428171781)

[1.6.2 Trial Management Group (TMG) 8](#_Toc428171782)

[1.6.3 Interim analysis 8](#_Toc428171783)

[2 Data sources, protocol non-compliance and analysis populations 8](#_Toc428171784)

[2.1 Data sources 8](#_Toc428171785)

[2.2 Protocol non-compliances 8](#_Toc428171786)

[2.3 Analysis populations 9](#_Toc428171787)

[2.3.1 Intention to treat (ITT) 9](#_Toc428171788)

[2.3.2 Per Protocol set (PP) 9](#_Toc428171789)

[3 Outline of statistical analyses 9](#_Toc428171790)

[3.1 General considerations 9](#_Toc428171791)

[3.2 Demographics and baseline characteristics 9](#_Toc428171792)

[3.2.1 Definitions and data manipulation 9](#_Toc428171793)

[3.2.1.2 Secondary outcomes scoring algorithms 11](#_Toc428171794)

[3.2.2 Recruitment and data completeness 12](#_Toc428171795)

[3.3 Clinical outcomes 13](#_Toc428171796)

[3.3.1 Primary outcome 13](#_Toc428171797)

[3.3.1.1 Statistical model 13](#_Toc428171798)

[3.3.2 Key secondary and secondary outcomes 13](#_Toc428171799)

[3.3.3 Analysis of the decision quality questionnaires 13](#_Toc428171800)

[3.3.4 Other outcomes 14](#_Toc428171801)

[3.3.4.1 Survival analysis 14](#_Toc428171802)

[3.3.5 Process evaluation 14](#_Toc428171803)

[3.3.5.1 Consultations 14](#_Toc428171804)

[3.3.5.2 Interviews 14](#_Toc428171805)

[3.3.5.3 Questionnaires 14](#_Toc428171806)

[3.3.5.4 Integration of analyses 15](#_Toc428171807)

[3.3.6 Subgroup analysis 15](#_Toc428171808)

[3.3.7 Model diagnostics 15](#_Toc428171809)

[3.3.7.1 Statistical analysis 15](#_Toc428171810)

[3.3.7.2 Secondary outcomes 15](#_Toc428171811)

[3.3.7.3 Analysis of the decision quality questionnaires 15](#_Toc428171812)

[3.3.7.4 Survival analysis 16](#_Toc428171813)

[3.4 Safety outcomes 16](#_Toc428171814)

[3.5 Missing and spurious data 16](#_Toc428171815)

[4 References 18](#_Toc428171816)

**List of abbreviations**

AE Adverse event

AFT Accelerated failure time

ANCOVA Analysis of covariance

CI Chief Investigator or confidence interval

CRF Case Report Form

CTRU Clinical Trials Research Unit

DESI Decision Support Instrument

DMEC Data Monitoring and Ethics Committee

EORTC European Organisation for Research and the Treatment of Cancer

ER Oestrogen receptor

GCP Good Clinical Practice

GEE Generalised estimating equation

GLM Generalised linear model

ICH International Conference on Harmonisation

IPQ Brief Illness Perceptions Questionnaire

ITT Intention to treat

PET Primary Endocrine Therapy

PP Per protocol

QoL Quality of life

RCT Randomised controlled trial

SAE Serious adverse event

SAP Statistical analysis plan

SD Standard deviation

SDV Source data verification

SOP Standard operating procedure

STAI State Anxiety Inventory

TMG Trial Management Group

# 1 Introduction, study design and key trial objectives

## 1.1 Study outline

The second phase of the Bridging the Age Gap study is a multi-centre, parallel group, pragmatic cluster randomised controlled trial (RCT) to evaluate whether use of a package of decision support interventions (DESIs), given to 50% of existing sites and embedded as ‘standard care’, helps to improve the quality of life (QoL), decision quality, decision regret, satisfaction and treatment understanding of older women entering the Age Gap study.

The aims of the study are:

1. To enhance the level of patient participation and decision quality in the treatment decision making process for older women with breast cancer. This will be achieved by making available bespoke and personalised patient facing DESIs focused on the following two choices:
   1. Use of surgery versus primary endocrine therapy (PET) in frailer less fit older women with oestrogen receptor positive (ER+) cancers.
   2. Use of adjuvant chemotherapy or no chemotherapy in older women with high risk cancers.
2. To improve and standardise the management of breast cancer in older women by use of a specially developed, evidence based, clinician facing management algorithm.

The study objectives are:

1. To assess the effectiveness of bespoke patient facing DESIs (PET versus surgery and chemotherapy versus no chemotherapy) in clinical practice in terms of improving patient QoL, decision quality (integrating knowledge, attitudes and decision made), coping and reducing decision regret, thus indicating better informed decision making.
2. To determine if, how, or to what extent, the interventions have an impact on clinical decision making among clinicians (change in PET/surgery rates and chemotherapy rates).
3. To determine whether the intervention improves patient knowledge and satisfaction with their preferred level of decision making.
4. To determine whether the intervention is effective in improving short, medium and long term cancer outcomes in this age group of women, (treatment morbidity, QoL, overall and disease specific survival).
5. To assess the utility and uptake of the package of DESIs from the perspective of both clinicians and patients. To this end a detailed process evaluation to primarily assess the usage of the DESIs (fidelity to the trial protocol), their usefulness and acceptability, and consider the facilitators and barriers to embedding them into everyday clinical practice.

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].

Running alongside this RCT will be a comprehensive process evaluation of the uptake and utility of the DESIs.

## 1.2 Hypotheses

The hypotheses of the trial are:

1. Use of clinical and patient facing DESIs will improve the QoL in older women with operable breast cancer and ultimately improve cancer outcomes.
2. Older women faced with a choice of treatment decisions for their breast cancer will report an improved decision quality and shared decision-making experience and less decision regret using DESIs compared to older women who receive usual standard clinical decision making support.
3. Use of evidence based DESIs will improve short and longer term outcomes by improving and standardising the quality of decision making, reducing the heterogeneity of practice across the UK.
4. Women in the intervention sites will express more positive illness representations (e.g. increased personal control, positive emotional consequences, less overall threat) and increased use of engagement coping strategies compared to women from the control sites.

## 1.3 Outcome measures

### 1.3.1 Primary endpoints

The primary endpoint is the global health status/QoL domain of the EORTC-QLQ-C30 [4], a generic QoL tool, measured at 6 months post-intervention.

### 1.3.2 Key secondary endpoints

The following QoL measures are key secondary endpoints (measured at 6 months post-baseline):

1. The functional and symptom scales of the EORTC-QLQ-C30,
2. The EORTC-QLQ-BR23 [5], a breast specific module, and
3. The EORTC-QLQ-ELD15 [6], an older person specific module.

***These are already collected as standard in the Age Gap cohort study.***

### 1.3.2 Secondary endpoints

The following secondary endpoints are new to the RCT:

Note: the timing of these questionnaires is around the treatment decision rather than entry to the study, so if baseline is after the treatment decision ( as it will be for the PET versus surgery decision, but not for the chemotherapy decision which will take place many months after baseline), then 6 weeks and 6 month measures are collected 6 weeks and 6 months after the treatment decision respectively.

1. Decision regret at 6 weeks and 6 months, measured by the Decision Regret Scale [7].
2. Shared decision making using CollaboRATE, at baseline [8].
3. Patient anxiety levels using the Spielberger short form State Anxiety Inventory (STAI) at 6 weeks and 6 months [9].
4. Knowledge, readiness to decide and treatment preference, measured using a non-validated questionnaire, at baseline.
5. Illness perceptions at 6 weeks and 6 months, measured using the Brief Illness Perceptions Questionnaire (IPQ) [10].
6. Coping at 6 weeks and 6 months, measured using the Brief COPE [11].

### 1.3.3 Other outcomes

The study will also compare survival outcomes and treatment types between control and intervention sites using data that is already collected via the normal cohort study protocol.

### 1.3.4 Process evaluation measures

The following process evaluation measures will be recorded:

1. DESI usage by trial staff and patients (staff and patient interviews, patient questionnaires and treatment decision CRF).
2. DESI acceptability (staff and patient interviews).
3. DESI usefulness (staff and patient interviews and patient questionnaires).
4. Barriers and facilitators to embedding the DESI into everyday routine clinical practice (staff interviews).
5. Levels of shared decision making determined by scoring of audio-recordings of consultations, patient interviews and CollaboRATE.

A selection of breast units randomised to the intervention arm (8 units) and a selection of control arm breast units (8 units) will be used for the process evaluation. These will be selected to be representative of the randomised units. Interviews with clinicians will only be conducted within the intervention arm sites and consultations will only be recorded when the chemotherapy decision is being discussed.

## 1.4 Sample size

The primary endpoint will be the global health status/QoL scale (questions 29 and 30 of the EORTC-QLQ-C30) at 6 months post intervention. Assuming 50 units are randomised to either the DESI intervention (25 units) or control (usual care - 25 units) then we can estimate a preliminary sample size assuming a fixed number of clusters (k=50) and try to recruit a set number of women per cluster [12]. Data from the EORTC Reference Manual [13] suggests a mean global health status/QoL scale of 58.2 with a standard deviation (SD) of 25.6 for women aged 70 or more with breast cancer. Cocks and colleagues [14] suggested the following guidelines for interpretation of the global health status/QoL that estimates for trivial, small, and medium mean differences were 1, 7, and 13 points respectively.

Assuming a SD of 26 points for the global health status/QoL scale and a mean difference of 7 or more points on the global health status/QoL scale between the groups is of clinical/practical importance (a “small” standardised effect size of 0.27). With no allowance for clustering; for the PET versus surgery DESI comparison with 291 eligible women per group we will have a 90% power of detecting this difference or more as statistically significant between the groups at the 5% two-sided level. If we assume an intra-class correlation of 0.05 then allowing for the clustered RCT design we will need to recruit 25 more women, eligible for using the decision aids, per cluster (i.e. 50 clusters x 25 women), 1,250 in total (this assumes a design effect of 2.2). With a 20% loss to follow-up by 6 months we need to recruit 34 women per cluster (50 clusters x 34 women) or 1,700 in total (850 per group). Based on our site recruitment data the majority of the sites will pass this number of cases after being in operation for 24 months.

## 1.5 Randomisation

Centres will be cluster randomised to one of two arms, stratified by high and low PET and high and low chemotherapy rates. Data for the stratification will be derived from the cohort study which has collected accurate data on treatment rates for both PET versus surgery and chemotherapy versus no chemotherapy. The two arms are:

1. **Control.** Usual standard practice for older women (>70 years) diagnosed with breast cancer with no change to normal counselling and decision making practice.
2. **Intervention.** Usual standard practice for older women (>70 years) diagnosed with breast cancer plus optional clinician and patient access to the DESIs (option grid, detailed information booklet and clinical algorithm) which will have been made available to these units to adopt as their standard of care.

## 1.6 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.

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

The DMEC will be composed of the following independent members: a geriatrician, a medical oncologist and a breast surgeon. The DMEC for the Age Gap study is already well established and comprises Professor Alistair Thompson, Professor Margot Gosney and Dr Matthew Hatton. They meet every 6 months and review study progress and a full DMEC report is produced. So far there have been no issues related to study progress. The study will continue with the same monitoring agreement.

### 1.6.2 Trial Management Group (TMG)

The TMG provides overall supervision of the trial, in particular: trial progress, adherence to the protocol, patient safety, input into the SAP and consideration of new information.

### 1.6.3 Interim analysis

There are no planned early stopping rules for this trial other than failure to recruit at a viable rate. This will be monitored by the TMG and DMEC. The study may be stopped after interim analysis after 12 months if the study is not meeting recruitment targets. This decision will be made by the TMG on the basis of advice from the DMEC.

# 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

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

For the purposes of the analyses, sites that are randomised to the intervention (DESIs) who do not implement/offer any part of the intervention will be classified as protocol non-compliances.

## 2.3 Analysis populations

### 2.3.1 Intention to treat (ITT)

This includes all patients for whom consent is obtained including those at protocol non-compliant sites (e.g. don’t want to use DESI). The ITT set will be used as the primary set for analysis and any other sets for sensitivity analysis.

### 2.3.2 Per Protocol set (PP)

This is a subset of the ITT set and excludes protocol non-compliant sites. Patients at intervention sites that choose not to receive the intervention will be included in the per protocol analysis, as this is a pragmatic study.

# 3 Outline of statistical analyses

## 3.1 General considerations

As the trial is a parallel group cluster RCT, data will be reported according to the CONSORT statement for cluster trials [15]. All statistical analysis exploratory tests will be two-tailed at the 5% significance level.

## 3.2 Demographics and baseline characteristics

Baseline socio-demographic (age, ethnicity), physical measurements and health related QoL data (EORTC-QLQ-C30, BR23, ELD15) will be summarised and assessed for comparability between the intervention and control groups. 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) and ethnicity, the number and percentage of participants in each of the categories will be presented.

All baseline summaries will be presented and reported for each group (DESI; control) and in total. 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 QoL scoring algorithms**

Scoring of the three questionnaires will be integrated into the database. The results will be checked using the raw data by the study statistician. For the EORTC-QLQ-C30 and EORTC-QLQ-BR23, the R package QoLR [16] will be used to calculate the scores. This package follows *The EORTC-C30 Scoring Manual (3rd Edition)* [17] for both scoring and dealing with missing data*.* For the EORTC-QLQ-ELD15, R functions will be written in order to calculate the scores.

EORTC-QLQ-C30 and EORTC-QLQ-BR23

The EORTC-QLQ-C30 and the EORTC-QLQ-BR23 will be scored according to *The EORTC-C30 Scoring Manual (3rd Edition)* [17], 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**

The EORTC-QLQ-BR23 comprises 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 [17] also discusses the method for dealing with missing data, which is described in Section 3.5.

EORTC-QLQ-ELD15

The EORTC-QLQ-ELD15 uses the same scoring system as for the EORTC-QLQ-C30 and the EORTC-QLQ-BR23 [17]. 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 [18], 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 [17], to report QoL on a scale of 0-100 where 100 indicates better QoL [6]. Therefore, mobility, joint stiffness, worries about others, future worries and burden of illness will be scored using the functional scale as defined in [17], whilst family support and maintaining purpose will be scored using the symptom scale as defined in [17].

### 3.2.1.2 Secondary outcomes scoring algorithms

Decision Regret Scale

The Decision Regret Scale consists of five items rated from 1 to 5. The process for calculating the score is as follows [19]:

1. Reverse code items 2 and 4 so that a higher value corresponds to more regret.
2. Convert each item to a 0 to 100 scale by subtracting one and multiplying by 25.
3. Sum the converted scores and take the average. A final score of 0 means no regret whilst a score of 100 means high regret.

CollaboRATE

CollaboRATE [8] consists of three questions rated from 0 to 9. Two scores can be built: CollaboRATE mean and CollaboRATE top score. For CollaboRATE mean, sum the three ratings and then multiply by 3.704 to get a final score on a scale of 0 to 100. A score of 0 means no effort was made whilst a score of 100 means every effort was made.

On the original publication the score was computed anyway in case of at least one missing item. However, as stated in the original publication, the acceptability of CollaboRATE items was demonstrated by less than 1% (8/1341) of participants missing any of the items.

For CollaboRATE top score For CollaboRATE top score participants should be coded as 1 (yes) when they recorded the highest response on the scale for all 3 items and as 0 (no)

in all other situations.

Short Form Spielberger State Anxiety Inventory (STAI)

The STAI six items (short form) consists of six statements measured on a 4-point Likert scale and will be scored as for the full STAI. After reversing the scores for positive statements, total scores are calculated (ranging from 6 to 24). A higher score therefore indicates greater anxiety. In case of a missing item the overall score is set to missing

Breast Cancer Treatment Choices

Correct responses are given a score of 1 whilst incorrect and ‘unsure’ responses are scored as 0. An overall score will be computed for sections A1 (range 0-8) and A2 (range 0-6) summing up the score of the single items.

Brief Illness Perceptions Questionnaire (IPQ)

Each item of the Brief IPQ assesses one dimension of illness perceptions. These are as follows:

|  |  |
| --- | --- |
| **Dimension** | **Question** |
| Consequences | 1 |
| Timeline | 2 |
| Personal control | 3 |
| Treatment control | 4 |
| Identity | 5 |
| Illness concern | 6 |
| Coherence | 7 |
| Emotional representation | 8 |

An overall score will also be computed by reversing the scoring for the personal control (3), treatment control (4) and coherence (7) and then summing to items 1, 2, 5, 6, and 8. A higher score reflects a more threatening view of breast cancer. Before using this overall score the internal consistency will be checked.

Brief COPE

The Brief COPE consists of 28 items made up of 14 two item scales. Eight of the scales are included in this study (i.e. 16 items overall), as shown below:

|  |  |
| --- | --- |
|  | **Item** |
| Active coping | 1, 5 |
| Denial | 2, 6 |
| Use of emotional support | 3, 10 |
| Use of instrumental support | 7, 14 |
| Behavioural disengagement | 4, 11 |
| Positive reframing | 8, 16 |
| Planning | 9, 12 |
| Acceptance | 13, 15 |

### Total scores on each scale range from 2 (minimum) to 8 (maximum). Higher scores indicate increased utilization of that specific coping strategy.In case of a missing item on the scale the score is set to missing.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. This information will be made available to the TMG and DMEC on request and as regular reports. The following will also be reported:

1. Number of patients screened per month.
2. Number of patients recruited per month.
3. Number and percentage of patients who complete each follow up or are lost to follow up.
4. 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 Primary outcome

The primary outcome will be the global health status/QoL domain of the EORTC-QLQ-C30 at 6 months post-intervention.

### 3.3.1.1 Statistical model

A marginal Generalised Linear Model (GLM), with coefficients estimated using generalised estimating equations (GEE) with robust standard errors and an exchangeable auto correlation matrix will be used to analyse the outcomes and allow for the clustered nature of the data. The exchangeable correlation structure corresponds to an equal correlation mode, meaning that the correlations of outcomes within a cluster, i.e. breast centres, are constant. For continuous outcomes, an identity link with a Normal distribution for the outcome will be used. Estimates for the intervention group coefficient from this regression model will be reported along with their associated 95% confidence interval (CI). In the event of differences between the intervention and control groups with respect to baseline demographic, physical, and health-related QoL measurements, then these covariates will be used in the GLM to adjust the intervention effect for these variables. The adjusted regression coefficient estimate for the intervention group parameter along with its 95% CI will then be reported. The Intracluster correlation coefficient will be reported as well.

### 3.3.2 Key secondary and secondary outcomes

For the key secondary and secondary outcomes, mean values for each domain of the questionnaires will be compared between the intervention and control groups, using similar models to those of Section 3.3.1.1. That is, where a questionnaire has more than one domain (e.g. EORTC-QLQ-C30, EORTC-QLQ-BR23, EORTC-QLQ-ELD15), the mean values for each domain will be compared between the intervention and control groups. Where a questionnaire has one outcome measure (e.g. the Decision Regret Scale), the mean value of the outcome will be compared between intervention and control groups. This analysis will be repeated for the 6 week and 6 month time-points (where applicable).

### 3.3.3 Analysis of the decision quality questionnaires

Attrition analyses will first be conducted to examine the characteristics of questionnaire non-responders at each follow-up, using chi-square and independent t-tests as appropriate. The primary analysis will use analysis of covariance (ANCOVA) to compare decision quality (measured using CollaboRATE) and decision regret (measured using the Decision Regret Scale) in women receiving the intervention versus usual care, controlling for baseline score (where appropriate) and other potential confounding factors.

Secondary analyses will involve using linear regression to examine the predictors (e.g. trial allocation, anxiety, illness perceptions and coping strategies) of decision quality (measured using CollaboRATE) and regret (measured using the Decision Regret Scale). The psychometric properties of the knowledge and collaborate components of decision quality will be examined using principal components analysis (underlying factor structure) and correlational analysis (convergence with other decision and coping measures).

### 3.3.4 Other outcomes

The following summary statistics will be reported overall and by intervention group (DESI intervention or control - usual care):

1. Number of deaths.
2. Number of deaths by cause (breast specific or other).
3. Number receiving PET, surgery, chemotherapy or no chemotherapy.
4. Overall survival at 6 months.

### 3.3.4.1 Survival analysis

Kaplan-Meier curves will be derived for each intervention group (DESI or usual care). Overall survival (OS) curves will be calculated using the Kaplan-Meier method. OS will primarily be compared between intervention groups using multivariate modelling, Cox’s Proportion Hazards model if appropriate. The Cox model should allow for frailty for taking in account the cluster structure of the trial. OS will also be compared using multivariate modelling adjusting for important prognostic factors. Hazard ratios (HR) and corresponding 95% CIs will be presented.

### 3.3.5 Process evaluation

The process evaluation will be completed by the team at Cardiff and consists of three parts: consultations, interviews and questionnaires. Data from the treatment decision CRF as well as details of the training workshops (to measure dose of training) will also be used for the process evaluation report.

### 3.3.5.1 Consultations

Audio recordings of the consultation will be analysed using the MAPPIN’SDM scale [20]. The MAPPIN’SDM scale measures shared decision making behaviours of the patient, the clinician and the combination of the patient and clinician (the patient-clinician dyad).

### 3.3.5.2 Interviews

Interview recordings will be transcribed verbatim. Analysis of patient and clinician interviews will initially be conducted separately. Thematic analysis will then be used to identify key themes within the sample. These will be guided by, but not exclusive to, the topic areas covered within the interviews (such as use of the DESIs, shared decision making, coping, barriers and facilitators to the DESIs, possible improvements) using a framework approach [21]. Analysis will be conducted by thorough reading of the transcripts and, using both deductive (guided by topic areas) and inductive (led by the data), key themes will be identified for coding. Within these codes, similarities and differences will be identified between transcripts in order to describe the content of each theme. Themes will also be examined according to high and low MAPPIN’SDM dyad scores (i.e. more/less shared decision making) for those using the chemotherapy DESI. Comparisons will also be made between the themes identified within the clinician and patient analyses.

### 3.3.5.3 Questionnaires

Questionnaires will be analysed using SPSS. Descriptive statistics will be used to report DESI usage within the intervention arm and by breast unit (based on questionnaire reports and CRFs). Descriptive statistics of CollaboRATE scores and MAPPIN’SDM scores (patient, clinician and dyad) will be calculated to give an indication of shared decision making. The associations between patient CollaboRATE scores and MAPPIN’SDM scores will be calculated. This will provide an assessment of agreement between actual and perceived levels of shared decision making.

Patient CollaboRATE scores will be compared between trial allocation groups, followed by subgroup analysis within each treatment option group (PET or surgery and chemotherapy or no chemotherapy). Mean differences will be assessed using Mann-Whitney U tests.

### 3.3.5.4 Integration of analyses

Results from the consultations, interviews and questionnaires will be integrated into one report to form the process evaluation. The Normalization Process Theory [22] will be used as a framework for the interpretation of how the DESIs have been implemented within the trial sites. Using the report, areas of improvement for the DESI will be identified (such as content and delivery of the DESIs and training needs) and then used to fine-tune the intervention and/or its implementation by making amendments. The results of the process evaluation will also be used alongside the overall trial results to try to identify which aspects of the DESIs did/did not work. Where differences between the trial allocation groups are shown in the main trial, it may be possible to use the process evaluation results to identify the potential active ingredient(s) in the intervention. The knowledge gained from the process evaluation will also be used to help with the interpretation of the trial results. In particular, fidelity to the trial protocol, levels of information provision and levels of perceived versus actual shared decision making in both trial arms will be important to consider alongside the effects (either significant or non-significant) of trial allocation on the main trial outcomes.

### 3.3.6 Subgroup analysis

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. severe) will be carried out on the primary endpoint only (global health status/QoL of the EORTC-QLQ-C30 at 6 months).

A plot of mean score for the global health status/QoL domain of the EORTC-QLQ-C30 at 6 months against subgroup will be produced. A subgroup by intervention interaction term will be fitted to the marginal GLM of Section 3.3.1.1 to test for a significant difference in outcome between subgroups. That is to say, the model will include intervention group and subgroup, and the interaction term between intervention and subgroup will then be added to test for a significant difference. We will test for subgroup effects even if the main intervention effect is not significant.

### 3.3.7 Model diagnostics

To check the fit of the marginal GLM, plots of the residuals against predicted values will be produced. Cluster Cook’s distance will be calculated for each participant to identify outliers. Any outliers will be removed and the model re-fit as a sensitivity analysis. Half-normal plots with simulated envelopes may also be considered to investigate goodness of fit.

## 3.4 Safety outcomes

Adverse events (AEs) and serious adverse events (SAEs) will be reported as number and percentage of patients overall and compared between intervention and control groups but no formal statistical analysis is planned.

## 3.5 Missing and spurious data

Any spurious data will be queried and checked for consistency with data management before data lock. Imputation methods will only be considered for the primary outcome (global health status/QoL of the EORTC-QLQ-C30 at 6 months). The imputation methods considered are as follows:

1. Multiple imputation. Baseline covariates (such as age, co-morbidity, treatment) will be among predictors of the missing data. Twenty datasets will be imputed.

This will be only performed for primary analysis.

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)* [17]. In summary, the following 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.

# 4 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 in Breast Cancer: Improving outcomes for older women. Evaluation of a decision support intervention for older women with operable breast cancer. A study nested within the Age Gap Cohort Study. Version 1, 4 May 2015.

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

[4] Aaronson, N.K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N.J., et al. The European Organization 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(5): 365-376.

[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] Brehaut, J.C., O’Connor, A.M., Wood, T.J., Hack, T.F., Siminoff, L., Gordon, E., et al. Validation of a decision regret scale. *Medical Decision Making*: 2003; 23(4): 281-292.

[8] Barr, P.J., Thompson, R., Walsh, T., Grande, S.W., Ozanne, E.M., Elwyn, G. The psychometric properties of CollaboRATE: a fast and frugal patient-reported measure of the shared decision-making process. *Journal of Medical Internet Research*: 2014; 16(1): e2.

[9] Marteau, T.M., Bekker, H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*: 1992; 31(Pt 3): 301-306.

[10] Broadbent, E., Petrie, K.J., Main, J., Weinmann, J. The brief illness perception questionnaire. *Journal of Psychosomatic Research*: 2006; 60(6): 631-637.

[11] Carver, C. S.  (1997).  You want to measure coping but your protocol’s too long:  Consider the Brief COPE. *International Journal of Behavioral Medicine*, 4, 92-100.

[12] Campbell, M.J., Walters S.J How to design, analyse and report cluster randomised trials in medicine and health related research. *Chichester: Wiley*. 2014.

[13] Scott, N.W., F Pea. EORTC QLQ-C30 reference values. EORTC, Brussels. 2008.

[14] Cocks, K., King, M.T., Velikova, G., Martyn St-James, M., Fayers, P.M., Brown, J.M. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. *Journal of Clinical Oncology*: 2011; 19(1): 89-96.

[15] Schulz, K.F., Altman, D.G., Moher, D., Group, C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Annals of International Medicine*: 2010; 152(11): 726-732.

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

[17] 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.

[18] 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: 852-858.

[19] O’Connor, A.M. User Manual - Decision Regret Scale [document on the Internet]. Ottawa: Ottawa Hospital Research Institute. © 1996 [modified 2013, cited 2015 07 16]. Available from:

<http://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Regret_Scale.pdf>

[20] Kasper, J., Hoffman, F., Heesen, C., Kopke, S., Geiger, F. MAPPIN’SDM--the multifocal approach to sharing in shared decision making. *PLoS One*: 2012; 7(4): e34849.

[21] Ritchie, J., SL. Qualitative data analysis for applied policy research. In: The Qualitative Researcher’s Companion. Eds: Huberman, A.M., Miles, M.B. 2002; Sage Publications. (London).

[22] May, C. FT. Implementing, embedding and integration practices: an outline of normalisation process theory. *Sociology*. 2009; 43(3): 535-554.

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

[24] 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.