Statistical Analysis plan for HYPOfractionated Radiation Therapy comparing a standard radiotherapy schedule (over three weeks) with a novel one week schedule in Adjuvant breast cancer: An open-label randomised controlled study (HYPORT- Adjuvant)

Trial registration

To be obtained

Statistical Analysis Version Number

1. Dated 15/11/18 - Initial version with Protocol version 1.0 submitted to IRB TMC

Protocol Version

This document has been written based on the protocol information as contained the study protocol version 1.0 dated 15/11/18

SAP Revision History

Revision	Date	Key Changes	Protocol Version
1.0	15/11/18	Initial Version submitted to IRB	1.0
2.0	3/12/18	Revised version with addition of information on interim analysis timing	2.0

Roles and Responsibilities

The statistical analysis plan was developed by the following:

Name	**Role**	**Signature**
Debashree	Drafting and designing, sample size calculation	

Signature of Senior Statistician:

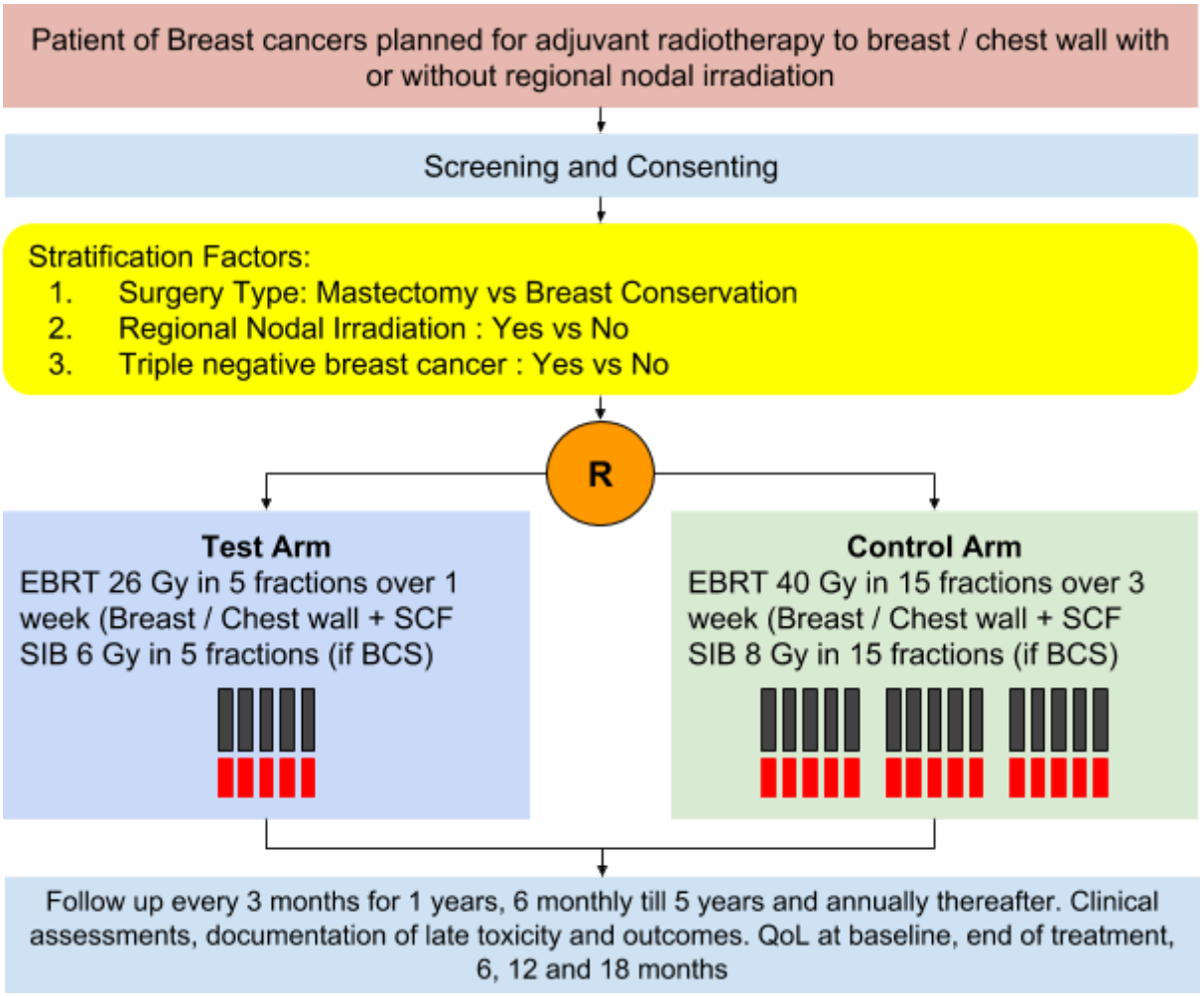
Signature of Principal Investigator:

Introduction

Background and Rationale

This randomized trial will compare the test intervention of a one-week hypofractionated radiotherapy course for adjuvant therapy for breast cancer against the standard three-week course of radiotherapy. Patients of breast cancer post mastectomy or breast conservation will be eligible for this study.

The trial schema is as below:



Trial Objectives

Research Hypothesis

The null hypothesis is that the one week course of radiotherapy is inferior to the standard

three-week course of radiotherapy, while the alternative hypothesis is that the one-week course is non-inferior to the three-week course of radiotherapy.

Study Objectives

Type	**Objective**	**Endpoint**
Primary Objective	To compare a one week course of hypofractionated radiotherapy to a three-week course on the cumulative probability of locoregional recurrence.	Cumulative probability of locoregional recurrence is the proportion of patients recurrence in the ipsilateral breast/chest wall or regional nodes after treatment.
Secondary Objectives	To compare a one week course of hypofractionated radiotherapy to a three-week course on invasive disease-free survival (iDFS)	Invasive Disease Free Survival (iDFS) is the time from randomization to the time of disease invasive recurrence, death due to any cause or any invasive second malignancy.
	To compare a one week course of hypofractionated radiotherapy to a three-week course on overall survival (OS)	Overall survival (OS) will be measured as the time from randomization to the time of death due to any cause.
	To compare the incidence of CTCAE Grade 3 or more late radiation-related adverse events in patients treated with one week course of hypofractionated radiotherapy to a three-week course (AE)	The incidence of occurrence of any Grade 3 or more radiation-related AE will be recorded.
	To compare the proportion of patients with quality of life (QoL) scores similar or better to the baseline scores between the two arms at 12 months	The proportion of patients with the 15 item EORTC QLQ C30 summary score similar to or better than the baseline scores in the two arms at 12 months.

Trial Methods

Trial design

The trial is designed as a prospective, two arm, parallel group, open-label non-inferiority randomized controlled trial with equal allocation to both groups. Patients will be randomized to either a one-week or three-week course of adjuvant radiotherapy after surgery for breast cancer.

Randomization

Stratified randomization using permuted blocks will be used for randomization. The randomization sequence will be generated using the Robust Randomization App (available at https://clinicalresearch-apps.shinyapps.io/rrapp/) and the randomization scheme generated from the app will be then used with the RedCap Randomization module to generate the randomization sequence. Patients will be allocated into the two groups in a 1:1 ratio. The following stratification factors will be used:

1. Type of Surgery: Mastectomy or Breast Conservation
2. Regional Nodal Radiation: Required or Not
3. Institution

Sample Size

The primary endpoint is locoregional recurrence rate and will be calculated using actuarial methods. For this endpoint, we assume that the locoregional recurrence rate in the control arm is 5% at 5 years and that the distribution is exponential in both arms. We hypothesize that the use of 1-week course of adjuvant radiotherapy will not increase the locoregional recurrence beyond 8% (absolute difference of 3%), corresponding to a hazard ratio of 1.63. We further assume that patients will be accrued over 5 years with an initial ramp up in the first year and that 2% of the patients will be lost to follow up each year of the trial. The total trial duration is expected to be 10 years to ensure a minimum follow up of 5 years for the last patient.

Framework

The current study will test for non-inferiority of the primary outcome i.e. locoregional recurrence rate. The secondary outcomes will be also be tested for non-inferiority.

Statistical Interim Analysis

The following interim analysis will be planned as a part of the study:

1. Interim Analysis 1: To demonstrate that acute adverse events in either arm do not exceed a pre-defined threshold.
2. Interim Analysis 2: To demonstrate that the trial intervention is not significantly inferior to the control.

Interim Analysis 1

This interim analysis will be performed after the first 100 patients have been accrued in the trial. During this analysis, the acute skin toxicity data of the patients will be compared against previously published outcomes. From the data on the acute toxicity of the FAST-Forward trial, we know that the risk of CTCAE Grade 3 or more acute toxicity is less than 3% in both the arms being investigated in the current study. The sample size of 50 patients would allow us to exclude a within-group rate of acute skin toxicity over 10% over the target rate of 5% with a power of 80% and a one-sided Type I error of 5%. Outcome data will not be evaluated during this interim analysis.

Interim Analysis 2

This will be used to evaluate if the one-week arm is significantly inferior to the three-week arm as a safety measure. The interim analysis for safety will be planned at 3 years when 16 events in total are expected to have occurred. If 16 events have not occurred then the interim analysis will be conducted at 3 years and guidance regarding continuation of the study will be obtained from the DSMC. As a result of this interim analysis, the sample size has been increased to 2100 patients. The critical Z value at the time of interim analysis is 3.47 for futility. The following table illustrates the key features of the design

Analysis	**Value**	**Efficacy**	**Futility**
IA1: 11%	Z	3.47	-1.48

N:1184	p(1-sided)	0.0003	0.9318
Events: 16	HR at bound	0.28	3.47
Year: 3	P(cross) if HR = 1.63	0.0003	0.07
	P(cross) if HR = 1	0.0056	0.008
Final	Z	1.96	1.96
N: 2100	p(1-sided)	0.025	0.025
Events: 140	HR at bound	1.17	1.17
Year: 10	P(cross) if HR = 1.63	0.025	0.975
	P(cross) if HR = 1	0.800	0.200

Asymmetric two-sided group sequential design with binding futility bound, 2 analyses, time-to-event outcome with sample size 2100 and 140 events required, 80% power, 2.5% (1-sided) Type I error to detect a hazard ratio of 1 with a null hypothesis hazard ratio of 1.63. Enrollment and total study durations are assumed to be 5 and 10 years, respectively. Efficacy bounds derived using a Lan-DeMets O'brien-Fleming approximation spending function. Futility bounds derived using a Lan-DeMets O'brien-Fleming approximation spending function.

The formal interim analyses will be performed by an independent Data Safety and Monitoring Committee which would be asked to give advice on whether the accumulated data from the trial, together with results from other relevant trials, justifies continuing recruitment of further patients or further follow-up. A decision to discontinue recruitment, in all patients or in selected subgroups will be made only if the result is likely to convince a broad range of clinicians including participants in the trial and the general clinical community.

Simulations for Alternative Event Rates

The following table illustrates the sample size requirements if control arm event rate is lower or higher than expected.

Control Arm 5 year	Experimental Arm 5	Hazard	Power	Type I	Sample

LRR	year LRR	Ratio		Error (one-sided)	Size
5%	8%	1.63	80%	2.5%	2100
4%	6.5%	1.62	80%	2.5%	2307
3%	4.85%	1.63	80%	2.5%	3069
6%	9.5%	1.61	80%	2.5%	1546
7%	11%	1.61	80%	2.5%	1328

Power simulation for secondary endpoints

The following are indicative results of the simulations done for power for the secondary endpoints:

1. **Overall Survival:** In the cohort of the patients used for calculation of the sample size is 92.8% (95% CI : 90.7% - 95%). The sample size of 2100 patients has a 77% power to exclude a 3% worsening of the overall survival rate from 93% in the control arm to 90% in the test arm assuming an one-sided type I error rate of 2.5%. A 3% change in overall survival is however unrealistic in this setting and more reasonable estimate of detriment in overall survival if the local control is worse by 3% is around 1%. Hence this study is underpowered to detect this difference as the sample size required to demonstrate a 1% non-inferiority would be 15,865 patients.
2. **Invasive Disease Free Survival:** The same cohort of patients, the invasive disease free survival including disease recurrence, second malignancies and deaths due to unrelated causes was 75% at 5 years (95% CI : 72% - 78.4%). The sample size of 2100 patients will allow us to exclude a 5% difference in the iDFS from 75% in the control arm to 70% in the test arm with 82% power and a one sided Type I error rate of 2.5%. This corresponds to a critical hazard ratio of 1.23.
3. **Adverse Events:** The estimated proportion of patients with Grade 2 or more late adverse effects in the standard arm is 11.4% for breast shrinkage and 2.8% for arm edema. The rate of arm edema applies to patients who receive regional nodal radiation therapy. In the current trial, the proportion of patients undergoing breast conservation is expected to roughly 60% of the total sample size. Also approximately 60% are expected to receive RT to the regional nodes. Thus we expect that approximately

1200 patients would be eligible for evaluation of these AE endpoints. The sample size of 1200 patients is adequate to detect a difference of 4% in the rate late breast shrinkage assuming that the rate is 11% in the control arm with a two sided type I error of 5% and a power of 80%. It is also adequate to detect a difference of 3.5% in the risk of arm edema assuming that the rate is 3% in the control arm with a two sided Type I error rate of 5% and power of 80%.

Final Analysis

The final analysis will be conducted once 140 events have occurred and it is expected that this will be performed at the 10th year of the trial. In case 140 events have not been observed even at 10 years, then the observation period will be extended for a maximum of 5 years or till 140 events occur.

Timing of Outcome Assessments

The schedule of the study procedures is indicated below. The start time for calculation of time to event endpoints will be the date of randomization. For the quality of life assessments, the timing of assessment is from the date of completion of radiotherapy. assessments

Assessment	All follow-up time points are taken from the date of randomization					
	Baseline	Adjuvant RT			3 - 12 months every 3 months	13 - 60 months every 6 months
		Wk 1	Wk 2	Wk 3		
Clinical Evaluation	X					
Eligibility Checklist	X					
Informed Consent	X					
Randomization	X					
RT QA	Prior to center initiation and throughout trial recruitment period					
RT Treatment (Control Arm)		X	X	X		
RT Treatment (Test Arm)		X				

RT Verification (Control Arm)		X	X	X		
RT Verification (Test Arm)		X				
SAE		X	X	X	X	X
Acute Toxicity		X	X	X		
FU Assessments				X	X	X
EORTC QLQ C30 (Control Arm)	X			X	At 6, 12 and 18 months after completion of RT	
FACT-B (Control Arm)	X			X	At 6, 12 and 18 months after completion of RT	
EORTC QLQ C30 (Experimental Arm)	X	X			At 6, 12 and 18 months after completion of RT	
FACT-B (Experimental Arm)	X	X			At 6, 12 and 18 months after completion of RT	

Statistical Principles

Confidence intervals, p-values and precision

For the primary as well as the secondary time to event endpoints, the applicable tests will be two-sided and will be performed using a 5% significance level. All confidence intervals presented will be 95% and will be two-sided. All values will be reported with a precision of 1 decimal place.

Adherence and Protocol Deviations

Compliance will be assessed based upon the percentage of patients completing the planned course of adjuvant radiotherapy. The percentage of patients not completing assessments at each time point for AEs and QoL will also be reported.

The following major protocol deviations will be reported:

1. Failure to complete radiotherapy within 1 week of the planned time.

2. Failure to adhere to the mandatory dose constraints for heart and lung.

Protocol deviations will be reported in the form of numbers as well as percentages of patients who will be summarised by treatment groups. While patients will be included in ITT analysis, they would be excluded from the per-protocol analysis for the time to event endpoints. The proportion of patients with any protocol deviations will be compared between the two groups using the Chi-square test.

Analysis Populations

The intention to treat population will include all randomized patients, regardless of their eligibility, according to the treatment they were randomized to receive. This will be applicable to all time-to-event endpoints. The safety population will consist of all randomised patients in this comparison who have received at least one dose of study treatment. Patients will be analysed according to the treatment they actually received. As this is a non-inferiority trial, the primary outcome will also be evaluated on a per-protocol basis to confirm the findings.

Trial Population

Screening Data

Screening data will be collected on an ongoing basis using a paper form entered into the RedCap database. The following information will be collected as a part of the screening process. The following screening information will be collected:

1. Total number of patients screened
2. Number of patients recruited for the trial
3. Number of screened patients not recruited
4. Reason for non-recruitment

An overall summary will be provided for this information.

Eligibility

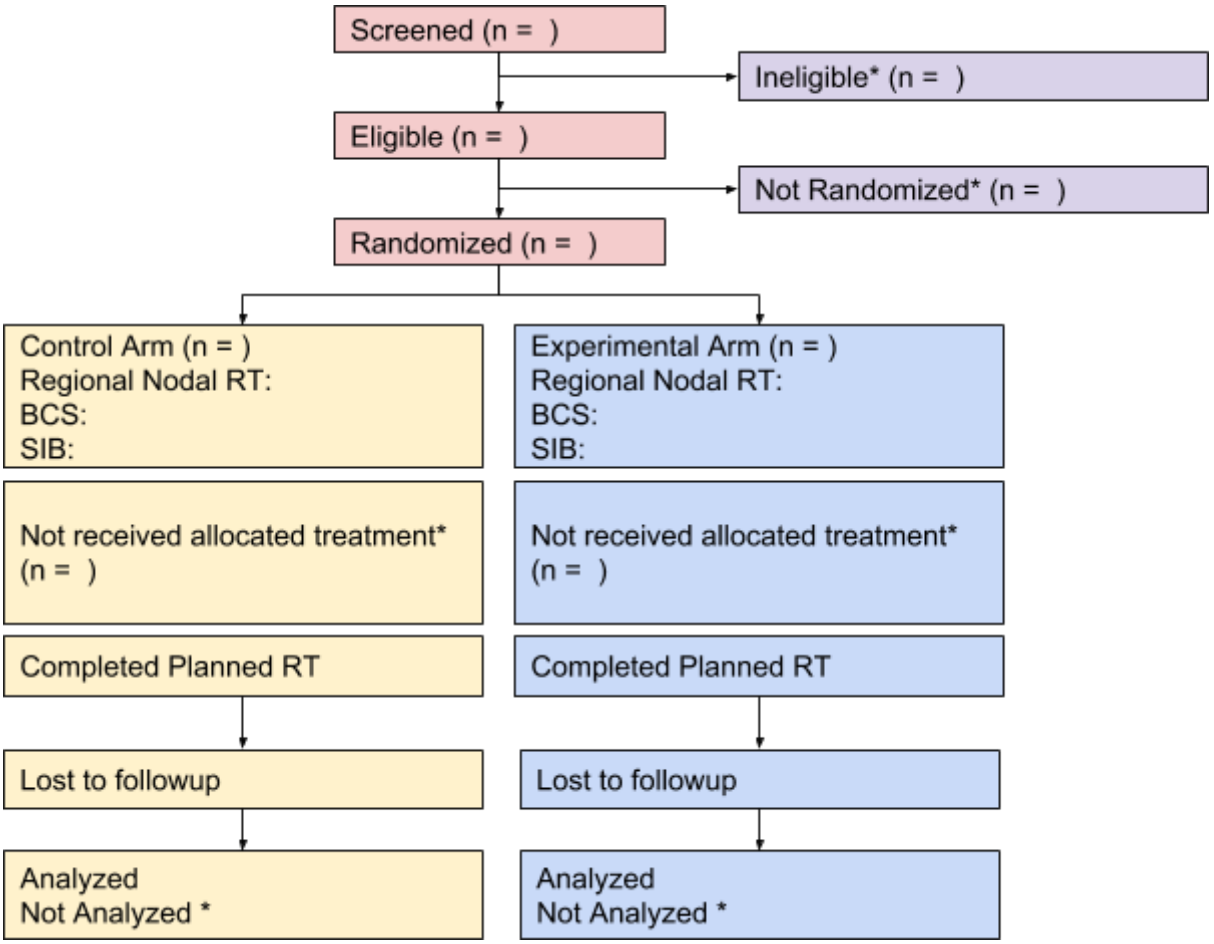
The details of the eligibility criteria are specified in the protocol document. The number of ineligible patients randomized if any, will be reported with reasons for ineligibility.

Recruitment

The CONSORT flow diagram to be included in this study will be used to summarize the number of patients who were:

1. Assessed for eligibility at screening
    a. Eligible at screening
    b. Ineligible at screening*
2. Eligible and randomized
3. Eligible but not randomized*
4. Received the randomised allocation
5. Did not receive the randomised allocation*
6. Lost to follow up*
7. Discontinued the intervention*
8. Randomized and included in the primary analysis
9. Randomized and excluded from the primary analysis*

For all items marked with an * reasons will be provided if applicable.



The figure above shows the template to be followed for the CONSORT flowchart for the trial. In boxes marked with * reasons will be provided with numbers.

Level of Withdrawal

The level of consent withdrawal will be tabulated (classified as "consent to continue to follow up and data collection", "consent to continue data collection only" and "complete-no further follow up or data collection").

Timing of Withdrawal

The timing of withdrawal will be presented a reverse Kaplan Meier plot of overall survival where patients who are alive will be considered to have events. This plot will have a number of risk table at intervals of one year and will be used to demonstrate the timing of withdrawal from follow up or loss to follow up data. In this plot, the number of patients who have died will be censored and a separate table below the plot will depict the censoring information for death in the same plot. The above plot will be incorporated with the CONSORT flowchart for

clarity.

The reason for losses to follow up (dropouts and withdrawals) over the course of the trial will be summarised by treatment arm and presented together in the main trial report. The reasons may include the following:

1. Patient discontinuation due to logistical reasons
2. Patient discontinuation due to death
3. Patient discontinuation due to unknown reason
4. Patient discontinuation due to toxicities
5. Patient discontinuation due to recurrence
6. Investigator deemed safety concern

Baseline Patient Characteristics

Patients will be described with respect to the age, menopausal status, performance status, surgery type (BCS vs Mastectomy), T stage, N Stage, Number of nodes dissected, Number of Nodes involved, grade, ER / PR / Her2neu status, lymphovascular invasion, perineural invasion, chemotherapy type (none, neoadjuvant, adjuvant, both), regimen (with or without taxanes), number of cycles of chemotherapy, and use of targeted agents.

Categorical data will be summarised by numbers and percentages, while continuous data will be summarised by median and interquartile range. Tests of statistical significance will not be undertaken for baseline characteristics and clinical importance of any imbalance will be noted.

Analysis

Outcome definitions

The following table lists the outcome definitions for the endpoints

Locoregional Recurrence	Any invasive recurrence in the ipsilateral breast or chest wall or ipsilateral axillary, supraclavicular or internal mammary lymph nodes (ipsilateral nodal lymph nodes level 1 - 4 and internal mammary nodes will be considered as an event. The time to event will be computed from the time of randomization. This will be calculated using the time to event method and will be estimated at 5

	years. The hazard ratio of locoregional recurrence between the two groups will be reported with 95% confidence intervals thereof. The estimate of the 5-year locoregional recurrence rate will be derived from cumulative incidence curves. See protocol for full definition.
Overall survival	The interval of time between the date of randomization to the date of death due to any cause. We will attempt to obtain the reason for death wherever feasible. Patients who are alive at the last follow up will be censored. The hazard ratio of overall survival between the two groups will be reported with 95% confidence intervals thereof. The estimate of the 5-year overall survival will be derived from Kaplan Meier curves.
Invasive Disease-Free Survival	The time from randomization to the time any recurrence (pre-invasive / invasive), distant metastases, death from any cause and second invasive primaries including invasive neoplasms of the breast. Patients who are alive and disease free at the last follow up will be censored. The hazard ratio of iDFS between the two groups will be reported with 95% confidence intervals thereof. The estimate of the 5-year iDFS will be derived from Kaplan Meier curves. As multiple mutually exclusive endpoints are included in this definition, the incidence of each event will be derived using the cumulative incidence method.
Adverse Events	The proportion of patients experiencing any Grade 3 or more adverse events will be calculated and compared between two arms.
Quality of Life	The 15 item summary score will be computed from the EORTC QLQ C30 questionnaire and this summary score will be used to calculate the proportion of patients with a score equal to or better than the baseline summary score. The proportion will be compared at 12 months.

Analysis Methods

For the primary endpoint locoregional recurrence as well as the secondary endpoints like overall survival and invasive disease-free survival, the hazard ratio and the 95% confidence intervals for the locoregional recurrence rate will be computed using the Cox regression analysis. The upper bound of the 95% confidence intervals would be used to demonstrate non-inferiority for locoregional recurrence which would be concluded if the same exceeds the predefined threshold of 1.63. The 5-year rate of locoregional recurrence rate will be graphically plotted using the cumulative incidence method. The analysis will be unadjusted for the covariates. The 5-year overall survival and invasive disease-free survival will be

graphically plotted using the Kaplan Meier method. As invasive disease-free survival comprises of multiple events, the rates of occurrence of the following events will be evaluated using the cumulative incidence method

1. Risk of distant metastases
2. Risk of second malignancies including breast cancer
3. Risk of distant metastases without locoregional recurrence

For the analysis of the adverse events, the proportion of patients with any Grade 3 or more adverse events will be calculated and will be compared between the groups using a Fisher's exact test. Additionally, we will also compare the following between the arms:

1. The proportion of patients with late toxicity Grade 2 or more between the two arms using a time to event methodology where the time will be calculated using the time from randomization to the time of the first appearance of the Grade 2 or more toxicity. If the patient has more than one Grade 2 or more toxicity events then these will be counted as one. The proportion will be compared using the Chi-square test.
2. The overall toxicity burden defined as the count of any CTCAE Grade 2 or more toxicities will also be compared between the two arms. Events occurring in the same patient or occurring more than once will be counted as seperate events. The total counts of toxicity will be divided by the number of patients treated per protocol in each arm to derive the mean number of toxicities per arm. This will be compared between the two arms using the Mann Whitney U test.
3. In addition to physician reported toxicity, patient reported toxicity outcomes will be obtained from the FACT B scores. The outcomes of interest will be questions related to breast cancer-specific subscales. For each patient we will compute a summary score comprising of the sum of all scores in the breast cancer subscale and the average of this score will be compared at 12 months using the Mann-Whitney test between the two groups.

Additional statistical analysis of toxicity in the form of logistic regression analysis may be undertaken to define the relationships between the toxicities and other clinical factors.

For the quality of life, a 15 item summary score will be calculated from the EORTC QLQ

C30. The proportion of patients who have a equal or better summary score from the baseline at 12 lines will be compared between the two groups. The comparison will be done using Chi-square test. Additional analyses to be done for QoL data will include:

1. Calculation of summary scores (mean and range) for all patients for the subscales of the EORTC QLQ C30 and FACT B questionnaires including the Trial Outcome Index (TOI). This will be done at baseline, end of therapy, at 6, 12 and 18 months. This will be depicted in the form of tables. Statistical comparisons will not be undertaken for this.
2. Graphical depiction of the QoL score trajectory over a period of time using the mean score as a metric.

Adjustment for Covariates

In addition to the unadjusted model which would be used for determining the non-inferiority, an adjusted model will be prepared for each time to event outcome of interest which would incorporate the following adjustment variables:

1. Age: As a continuous variable
2. Menopausal Status: Yes or no
3. Baseline ECOG PS: 0 -1 versus 2 - 3
4. Nodes involved: As a continuous variable
5. Grade: 1-2 versus 3
6. ER/PR Status: As a continuous variable using the Allred score
7. HER2neu status: As a categorical variable
8. Ki67 level: As a continuous variable.
9. Lymphovascular space invasion: Present or absent
10. Perineural Invasion: Present or absent
11. Relative chemotherapy dose intensity: As a continuous variable. Will be considered as 100% in patients who are not planned for chemotherapy.
12. Boost: Simultaneous Integrated Boost versus Sequential Boost

Additional adjustment variables may be chosen as dictated by the TMC / DMSC and the same will be incorporated in the SAP as a version modification. For all endpoints, the results of both adjusted and unadjusted analysis will be reported but the primary outcome will be

evaluated using the unadjusted analysis.

Model assumptions checks and alternative methods

The model diagnostic check for the Cox regression model will include checking for influential observations, linearity and proportional hazards assumptions. For this established diagnostic procedures which incorporate evaluation of various residuals will be used. Both statistical testing and graphical plotting will be used for this evaluation.

1. For testing the proportional hazards assumption the scaled Schoenfeld residuals will be plotted against the log of time. The statistical test will be testing for a non-zero slope in a generalized linear regression of the scaled Schoenfeld residuals on the functions of time. If the proportional hazards assumption is not met then the accelerated failure time (AFT) model will be used. If the fit of the AFT model is not adequate, as demonstrated by a Q-Q plot, then a restricted mean survival time method as demonstrated by Royston & Parmar will be utilized(1). The restricted mean survival time approach is robust as it does not assume proportional hazards.
2. Influential observations or outliers will be checked using deviance residuals which are normalized transformations of the martingale residuals. The plot of these residuals against the observations should symmetrically distribute about zero with a standard deviation of 1.
3. Linearity assumptions for continuous covariates will be checked by plotting the Martingale residuals against the covariate. Fitted lines with loess function should appear linear to satisfy the linearity assumptions. If the same is found to be violated, continuous variables will be expanded by restricted cubic splines with 3 or more knots will be used to model the continuous covariate.
4. In order to evaluate the model discrimination and calibration, bootstrapping with 500 bootstrap samples will be used to derive a calibration plot and optimism corrected C-index.

Sensitivity Analysis

While every effort will be made to avoid missing data, a degree of missingness is anticipated. The trial sample size is calculated to account for a 2% loss to followup per year of the trial.

However, if more than 10% of the population is lost to follow up then sensitivity analysis will be undertaken for the quality of life and adverse events endpoints.

Missing data for AEs and QoL will be analysed graphically stratified by arm and dropout time to evaluate if the missing data are Missing at Random or not. Sensitivity analysis will comprise of the following imputations to allow for MNAR conditions:

1. Best score is imputed for all
2. Worst score is imputed for all
3. Worst score imputed for the control arm and best score for the experimental arm
4. Best score imputed for the control arm and worst score for the experimental arm

The analyses will be repeated using these imputed datasets to generate sensitivity estimates of the QoL data. The above methodology provides a conservative estimate of the quality of life estimates. Additionally, another analysis will be conducted using complete cases only.

Subgroup Analysis

The following pre-specified subgroup analyses will be conducted on the primary and the other time to event endpoints stratified by:

1. Menopausal Status: Premenopausal vs Postmenopausal
2. Nodes Positive: None vs 1 - 3 vs > 3
3. Subtype: Luminal vs Her2-enriched vs TNBC
4. Regional Nodal Radiation: None vs SCF alone vs IMN + SCF
5. Boost Type: SIB vs Sequential Boost

These subgroups are chosen as they have an implication on the primary as well as the secondary time to event outcomes and are independently prognostic. Results will be presented as forest plots with interactions results alongside. The interaction test will test if the treatment effect is modified by the subgroup.

In order to evaluate the impact of prognostic factors on the treatment effect, a nomogram will be built using Cox regression that includes treatment as a factor. The use of this nomogram will enable the user to judge the benefit arising from the use of a particular fractionation

regimen in a patient with a given combination of prognostic factors.

Statistical Software

All statistical analysis will be conducted using R (R Core group, Vienna, Austria) or similiar software. Packages used for analysis will be quoted. The analysis will be conducted using a IDE like RStudio and the code of the analysis will be presented as a markdown file for evaluation.

Data Handling and Cleaning

Data for the trial will be stored on a custom designed RedCap database hosted at Tata Medical Center. The software provides rules for checking data completeness and validation. The data completeness will be checked at regular intervals by the TMC. The DSMC will have access to the data for the evaluation of the data during interim analysis and as mandated by the IRB and DSMC.

References

1. Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med Res Methodol. 2013 Dec 7;13:152.