Statistical Analysis Plan

Development Of A New Approach To Measure The Volume Of Breast Tissue Removed During Wide Local Excision Using Specimen Weight And Mammographic Density

Estimating breast specimen volume from weight and radiological density

	Name	Signature	Date *(DD/MMM/YYYY)*
Chief Investigator	Emanuele Garreffa		
Study Statistician	Jacqueline Beckhelling		

Protocol Version Number and Date	V1.0 24/Jun/2023	**IRAS reference**	325661

Version Number	Date	Author	Description of Changes
V1.0	04Mar2024	J Beckhelling	Original document

TABLE OF CONTENTS

[A] Background

[A.1.1] This is a proof-of-concept study of tissue removed from women undergoing surgery for breast cancer. It is intended to produce a novel method of estimating the volume of the tissue removed. The volume of tissue removed influences the patients' follow-up treatment and also the size of the implant that is used in reconstruction, so it is important to have an accurate measure or estimate of the volume. There is a method of measuring the breast tissue that has been removed (the Archimedes/water displacement method), however it is somewhat time-consuming and it is not feasible to use it in usual care due to the volume of operations that are required. There is an existing method of estimating the volume of the samples, which is to divide the weight (in grams) by a constant value (0.958). However, the estimates this produces vary in accuracy, possibly because they do not account for features of the tissue sample that will affect the relationship between its weight and its volume.

[B] Sample Size

[B.1] Sample Size Estimation / Power Calculation [Protocol Sec. 10.1]

[B.1.1] This study has two phases. Phase 1 will require tissue samples from 110 patients. In this phase the regression model that generates the most accurate estimates of excised tissue volume using data readily available to clinicians will be found using stepwise regression. A search of the literature found no data upon which a sample size calculation could be based for this phase. Therefore, the sample size for this phase of the data was based on a "rule of thumb" after reviewing Green[3].

[B.1.2] Phase 2 of this research will compare the accuracy of the estimates obtained from the regression model identified in phase 1, the estimates obtained by the current estimation method (i.e. dividing the weight by 0.958) and the gold standard measure of tissue volume (using the Archimedes/water displacement method). The accuracy of the estimates will be reported and assessed as percentages of the measured sample volume. The accuracies will be assessed using limits of agreement analyses and will require a sample size of 246 tissue samples. This sample size was calculated using a half width for the limits of agreement confidence interval of 5% (i.e the fill confidence interval will be LOA +/- 5%). This was chosen as a difference of more than 5% could affect the accuracy of the follow-up boost radiotherapy in the Chief Investigator's clinical judgement. The standard deviation of the differences using the current method was 23.26%, based on a clinical audit carried out by the CI between January and December 2022 which measured the tissue volumes using the water displacement/Archimedes method. The audit only included 50 patients although approx. 480 RDH patients receive this surgery per year due to the difficulty of measuring the tissue volumes as part of usual care.

[B.2] Sample Size Amendments After Interim Analysis

[B.2.1] Not applicable, there will be no interim analysis of either phase of the research.

[B.3] Final Sample Size

[B.3.1] Phase 1: 110 tissue samples, Phase 2: 246 tissue samples.

[C] Randomisation

[C.1.1] Not applicable.

[D] Interim Analysis [Protocol Sec. 10.6]

[D.1] Justification for Interim Analysis

[D.1.1] Not applicable, no interim analyses will be carried out in either phase of the research.

[D.2] Definition of Estimands used in Interim Analysis

[D.2.1] Not applicable.

[D.3] Statistical Methods for Interim Analysis

[D.3.1] Not applicable.

[E] Final Statistical Analysis

[E.1] Summary of Baseline Data [Protocol Sec. 10.3.1]

[E.1.1] Two baseline variables will be collected in both phases of the study, age and BMI. They will be reported using means and standard deviations.

[E.1.2] This is not a randomised clinical trial therefore a Consolidated Standards of Reporting Trials (CONSORT) flow diagram is not required.

[E.2] Definition of Primary Estimand or Endpoint [Protocol Sec. 3.2]

[E.2.1] Phase 1: A regression model that produces the most accurate estimates of tissue volumes using some or all of the following:

1. mass present (yes/no) - This will indicate if a tumour mass was observed in the sample.
2. mass diameter A (mm) - this will be the largest tumour diameter in the tumour plane with the greatest area (assessed by eye by the radiologist). It will be included as an interaction term with mass present as it will be missing if a tumour has not been observed.
3. mass diameter B (mm) - this will be 90^{o} to the plane of the tumour with the greatest area. It will be included as an interaction term with mass present as it will be missing if a tumour has not been observed.
4. healthy tissue density (based on the specimen x-ray density and assessed using a VAS) - this is the density of the healthy margin of tissue surrounding the tumour, which can vary from woman to woman. If no mass has been observed, this will be the density of the complete tissue sample.

5. area occupied by the tumour (percentage of specimen x-ray) - this could be important because the tumour will always be denser than the surrounding healthy tissue. It will be included as an interaction term with mass present as it will be missing if a tumour has not been observed.
6. specimen weight (g)
7. tumour histology - this affects the tumour shape. It will be included as an interaction term with mass present as it will be missing if a tumour has not been observed.
8. tumour grade - this measures how closely the tumour cells resemble healthy tissue. It will be included as an interaction term with mass present as it will be missing if a tumour has not been observed.
9. patient age - this is related to how dense the healthy breast tissue is.
10. patient BMI - this is related to the size of a woman's breasts, which affects the sample that is removed.

The dependent variable will be the tissue volumes measured using the Archimedes/water displacement method.

[E.2.2] Phase 2: the accuracy of the volume estimates obtained from the regression model and the current method (dividing the specimen weights by 0.958, regardless of their radiological density) will be compared with the gold standard measurement of specimen volume (the water displacement / Archimedes method).

[E.3] Statistical Methods for Primary Analysis [Protocol Sec. 10.3.2]

[E.3.1] Phase 1 - The regression model providing the most accurate estimates of tissue volume will be found using di-directional stepwise regression.

Mullticollinearity may be a concern for this dataset. The CI had collected a small sample (of 50 patients) of some of the variables we intend to collect in this study as part of an audit carried out between January and December 2022. In this dataset two radiologists assessed or measured six variables. The data below are the assessments and measurements made by one of the radiologists (=radiologist R1). These data were investigated to identify variables that were correlated, with the results found below:

NB SpecVol = specimen Volume

SpecWgt = specimen weight

R1DiamA = the largest diameter of the tumour (assessed by the radiologist by eye)

R1DiamB = the diameter at 90^{o} to the plane with the largest diameter/largest tumour area (assessed by the radiologist by eye)

R1Dens = the density of healthy tissue surrounding the tumour in the sample

R1Area = the tumour size (assessed as a percentage of the total sample size)

```
. pwcorr SpecVol SpecWgt R1DiamA R1DiamB R1Dens R1Area
```

	SpecVol	SpecWgt	R1DiamA	R1DiamB	R1Dens	R1Area
SpecVol	1.0000					
SpecWgt	0.9233	1.0000				
R1DiamA	0.0433	0.1025	1.0000			
R1DiamB	0.0854	0.1716	0.9339	1.0000		
R1Dens	-0.0022	-0.0599	0.0091	0.0277	1.0000	
R1Area	-0.1866	-0.1335	0.7446	0.7338	-0.0162	1.0000

It is not surprising that the diameters (R1DiamA and R1DiamB) are so highly correlated. However, 9 of these patients had no tumour (they had an early form of breast cancer called DCIS). As they had no tumour both diameters for these patients were set to zero. If these patients are removed from the dataset. The multicollinearity changes somewhat:

```
. *Getting a correlation matrix
. pwcorr SpecVol SpecWgt R1DiamA R1DiamB R1Dens R1Area
```

	SpecVol	SpecWgt	R1DiamA	R1DiamB	R1Dens	R1Area
SpecVol	1.0000					
SpecWgt	0.9123	1.0000				
R1DiamA	0.1934	0.2432	1.0000			
R1DiamB	0.2524	0.3384	0.8820	1.0000		
R1Dens	-0.0460	-0.1087	0.1564	0.1813	1.0000	
R1Area	-0.1716	-0.1143	0.6949	0.6765	0.0487	1.0000

```
.
end of do-file
```

The correlation between the specimen volume (the dependant variable) and all of the other variables is stronger which suggests the inclusion of the cases without tumours are masking the relationship between the tumour details and the specimen volume somewhat. However, with the exception of specimen weight, all of the correlations between the explanatory variables and the independent variable (tumour volume) are still weak (i.e. less than an absolute value of 0.4).

Among the explanatory variables, R1DiamA and R1DiamB have strong correlations with R1Area, as would be expected and R1DiamA has a strong relationship with R1DiamB, but the other variables only have weak correlations.

Multicollinearity will be considered prior to fitting the regression model, when data on all of the explanatory variables can be included. A possible strategy would be to fit two models. One including the area, but neither diameter and the other including one or possibly both of the mass diameters (combined together in some way) but not including the area. However, the final decision of how to manage multicollinearity will be made after all of the explanatory variables have been assessed and it will be reported in the Technical Verification Report.

Nine of the 50 samples available from the audit had no tumour (18%). If this is a typical finding, we should expect approximately 20 patients in the sample of 110 to have no tumour. Since these samples will have no data about tumour size, shape (which will be

collected in the study dataset) and area, the model will be specified to include R1mass (whether a tumour is present or not) as a main effect and the interactions of R1mass and any variable that will be missing if there is no tumour mass present (e.g. diameter A, diameter B, Area and shape). The variables dependent on the existence of a tumour mass will not be included as main effects in the model.

The variables to be included or removed from the regression models will be selected using the Akaike Information Criteria (AIC). This is preferable to using p-values, because the AIC includes a penalty for every variable added to the model, which reduces the risk of overfitting, which is not the case if p-values are used for variable selection. A similar criterion used for model selection is the Bayesian Information Criteria (BIC). The BIC will identify the correct/true model if the correct/true model is among the possible models that can be fitted from the variables that have been collected (Vrieze[4]). However, this model will be based on data readily available to clinicians which may not include all of the variables needed to identify the correct/true model. In contrast, the AIC aims to select the model that minimises the MSE of predictions. Since the model will be used to predict the tissue volumes, this feature of the AIC is particularly important and therefore the AIC is preferred over the BIC. For small sample sizes, a corrected version of AIC is available, AIC_c. Using AIC_c was considered, but a paper by Brewer et al.[2] simulated different datasets (e.g. with different degrees of correlations and heterogeneity within the model parameters) and assessed the root mean square error (RMSE) of the models fitted from different sized samples of these datasets using AIC, AIC_c and BIC. The plots of RMSE for sample sizes = 100 for AIC and AIC_c were the same or extremely similar, therefore AIC_c will not be needed for the variable selection for the model.

The stepwise regression will be fitted using bi-directional stepwise regression using the built-in procedure "step" in r. Variables will be included if they reduce the Akaike Information Criteria (AIC). They will be excluded if, after other variables have been included in the model, they no longer make an important improvement in the model's fit (based on the AIC). The model will stop fitting variables when, based on the AIC, none of the variables that are not already included will make an important difference to the model's fit if they are included and all of the variables that are included do make an important difference, so none of them should be removed. The best fitting model will be the model with the smallest value for AIC.

After the regression model has been fitted, the residuals will be analysed to assess normality and randomness. If the residuals are not normal and randomly distributed, variables will be transformed to improve the model.

[E.3.2] Phase 2: The accuracy of the regression model, and the current method will be compared to the gold standard (measurement using the Archimedes/water displacement method) using limits of agreement.

For phase 2 of the research, 95% limits of agreement (Bland and Altman[1]) with 95% confidence limits will be calculated comparing the novel method (using the regression model) and the current method (dividing the specimen weights by 0.958, regardless of their radiological density) to the gold standard measurement of specimen volume (the water displacement / Archimedes method). The data will be reported using Bland-Altman plots. As the current method is usual care, the acceptable limit of agreement is the limit of agreement for the current method.

Data from a second sample of patients, who did not supply information for phase 1 will be collected for phase 2.

[E.4] Definitions of Secondary Estimands or Endpoints [Protocol Sec. 3.2]

[E.4.1] Phase 1 - No secondary analyses will be carried out.

[E.4.2] Phase 2 - The similarity of the accuracy of estimates with each estimation method and the differences between the methods will be assessed.

[E.4.3] Phase 2 - The within patient differences between the two estimation methods and the gold standard will be compared.

[E.5] Statistical Methods for Secondary Analyses [Protocol Sec. 10.3.2]

[E.5.1] Phase 2 - The intraclass coefficient will be calculated to establish how similar the accuracy of the estimates is within each estimation method, and how different they are between the methods. The accuracy will be measured as a percentage because an absolute difference of 5ml on a tissue sample of 20ml would be a large difference, whereas a difference of 5ml for a 100ml would not. The data will be modelled using a two-way mixed effect model of patient, estimation method and estimation accuracy, with the two estimation methods treated as fixed effects. The intraclass correlations will be reported with 95% confidence intervals.

[E.5.2] Phase2 - The within-patient accuracy of the two methods will be assessed with a paired t-test. The differences between the tissue volume estimates from the regression model and the measured volumes (the gold standard) will be calculated. As will the differences between the current method and the measured volumes. The mean and SDs of the accuracy of the two methods will be reported and compared using a paired t-test.

Statistical Methods for Sub-group Analyses

[E.5.3] This is a proof-of-concept study, subgroup analyses will not be carried out.

[E.6] Statistical Methods for Sensitivity Analyses

[E.6.1] The stepwise regression will be re-fitted using the combined sample of women from phase 1 and phase2 (346 women). The purpose of this analysis is to assess to what degree the model changes if the sample size is increased. This will indicate whether the model should be refitted in later research when a sample size can be calculated for the model fitting.

[E.7] Definition of Safety Endpoints

[E.7.1] Adverse events are not expected in this study, as it investigates the best way to estimate the volumes of tissue excised from patients. However, should adverse events occur they will be reported.

[E.8] Statistical Methods for Safety Endpoints

[E.8.1] In the unlikely event that adverse events occur, they will be reported using frequencies and percentages.

[F] Analysis Groups and Missing Data

[F.1] Definition of Analysis Groups [Protocol Sec. 10.7]

[F.1.1] Only tissue samples with complete data will be included in the analyses.

[F.2] Procedure for Accounting for Missing, Unused, and Spurious Data [Protocol Sec. 10.8]

[F.2.1] Tissue samples with incomplete data will not be included in the analyses and will be replaced. Phase 1 requires 110 patients with complete data and phase 2 requires 246 samples with complete data.

[G] Unplanned Analyses

[G.1] Unplanned Analyses Requested by the CI

[G.2] Unplanned Analyses Requested by the Sponsor

[G.3] Unplanned Analyses Requested by the Journal Reviewer

[H] Comments

[H.1.1] Phase 1: the bi-directional stepwise regression will be carried out using the built-in "step" procedure in r. The residuals will be investigated using STATA. Phase2: all analyses will be carried out using STATA.

[H.1.2] References:

1. Bland JM, Altman DG, Measuring agreement in method comparison studies. Statistical Methods in Medical Research. 1999 vol 8(2):135-160. doi: 10.1177/096228029900800204.

2. Brewer MJ, Butler A, Cooksley SL, The relative performance of AIC, AIC_c and BIC in the presence of unobserved heterogeneity. Methods in Ecology and Evolution. 2016. Vol7:679-692. doi: 10.1111/2041-210X.12541

3. Green SB, How many subjects does it take to do a regression analysis. Multivariate behavioral research. 1991. Vol26(3):499-510. doi: 10.1207/s15327906mbr26037.

4. Vrieze SI, Model Selection and psychological theory: A discussion of the differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Psychol. Methods. 2012. Vol 17(2): 228-243. doi:10.1037/a0027127.

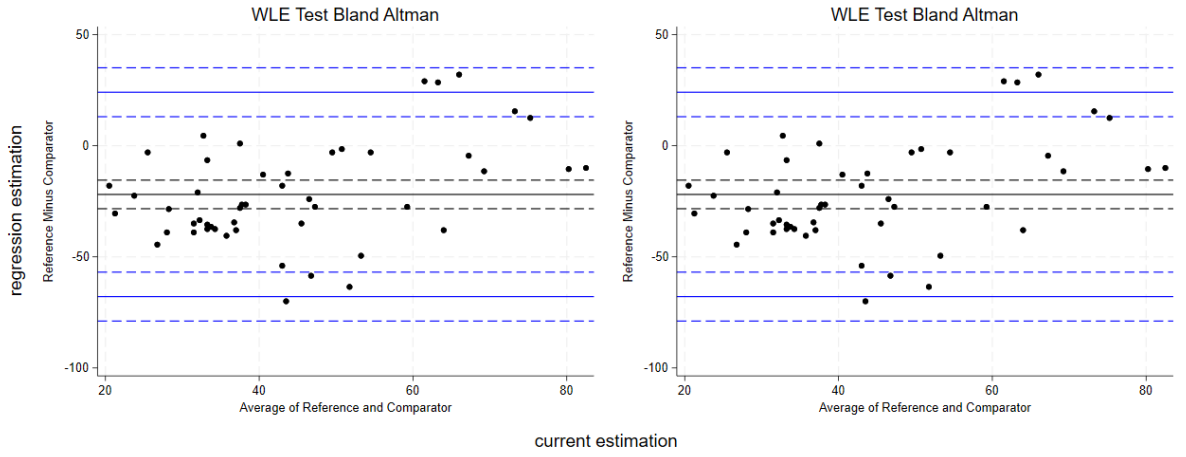
[I] Appendix A: Dummy Tables

[I.1] Table 1: Summary of Stepwise regression of 110 tissue samples

SAP section ref	**PHASE 1: PRIMARY ANALYSIS**	**Variable**	**Added or removed**	**Akaike Information Criteria**
D.3.1	**Step** **1 (intercept only)** **2** **3**	NA	NA	
	(steps will be added (or removed) so that all of the steps included in the stepwise procedure are shown in this table)			
	Regression model: SpecVol = Intercept +….			

[I.2] Figure1 : Bland-Altman Plots of current estimation method vs measurement of tissue volumes

NB the image below shows two images of the same graph. It is intended to show the format of the two graphs that will be produced, one showing the current estimation vs the measured volumes and the other the regression estimates vs the measured volumes.



Graph Key:

Solid black line	The bias (mean difference) between the two methods
Dashed black lines	The 95% confidence intervals for the bias
Solid blue lines	The 95% Limits of Agreement
Dashed blue lines	The 95% confidence intervals for the Limits of Agreement

[I.3] Table 2: Bland-Altman Measurements

SAP section ref	**PHASE 2: PRIMARY ANALYSIS Bland-Altman analyses**	**Bias (95% CI)**	**Upper 95% LOA* (95% CI)**	**Lower 95% LOA (95% CI)**
D.3.1	**Current estimation vs measured volume (ml)**	*X (X to X)*	*X (X to X)*	*X (X to X)*
	Regression estimation vs measured volume (ml)	*X (X to X)*	*X (X to X)*	*X (X to X)*

*LOA = Limit of agreement.

[I.4] Table 3: Intraclass correlation

SAP section ref	**PHASE 2: SECONDARY ANALYSIS: Intraclass Correlation**	**ICC**	**95% Confidence interval**	
			Lower bound	**Upper Bound**
	Individual	*X*	*X*	*X*
	Average	*X*	*X*	*X*

[I.5] Table 4: Paired t-test

SAP section ref	**PHASE 2: SECONDARY ANALYSIS: Paired t-test**	**Current estimation method** (n=*N*)	**New estimation method** (n=*N*)	**Difference between methods**	**p-value**
D.3.1	**Mean (95% CI)**	*x (x to x)*	*x (x to x)*	*X (X to X)*	*X*

[I.6] Table 5: Sensitivity analysis - Stepwise regression based on 346 tissue samples

SAP section ref	**SENSITIVITY ANALYSIS Summary of Stepwise Regression on all samples (n = 346)**	**Variable**	**Added or removed**	**Akaike Information Criteria**
D.3.1	**Step**			
	1 (intercept only)	NA	NA	
	2			
	3			

SAP section ref	**SENSITIVITY ANALYSIS** **Summary of Stepwise Regression on all samples (n = 346)**	**Variable**	**Added or removed**	**Akaike Information Criteria**
	(steps will be added (or removed) so that all of the steps included in the stepwise procedure are shown in this table)			
	Regression model: SpecVol = Intercept +….			

[I.7] Table 6: Demographic data

SAP section ref	**SUMMARY OF BASELINE DATA**	**Phase 1** (n=*N*)	**Phase 2** (n=*N*)	**All patients (n=N)**
D.1.1	**Age (years)** *mean (95% CI) or median (IQR)*	*X (X to X)*	*X (X to X)*	*X (X to X)*
	BMI *mean (95% CI) or median (IQR)*	*X (X to X)*	*X (X to X)*	*X (X to X)*

[I.8] Table 7: Adverse events (NB this may not be required)

SAP section ref	**SAFETY ENDPOINTS**	**Study Population** (n=*N*)
D.9.1	**AE diagnosis** *n (%)*	-
	Diagnosis A	*X (X)*
	Diagnosis B	*X (X)*

[I.9] Table 8: Number of Serious adverse event and adverse reactions (this may not be required)

SAP section ref	**SAFETY ENDPOINTS**	**Treatment X** (n=*N*)	**Treatment Y** (n=*N*)	**Difference between treatments**	**p-value**
D.9.2	**Number of SARs** *mean (95% CI) or median (IQR)*	*X (X to X)*	*X (X to X)*	*X (X to X)*	*X*
D.9.3	**Number of SAEs** *mean (95% CI) or median (IQR)*	*X (X to X)*	*X (X to X)*	*X (X to X)*	*X*