

Effectiveness of using texture analysis in evaluating heterogeneity in breast tumor and in predicting tumor aggressiveness in breast cancer patients

Alix Hopp (UACOM Phoenix MS4), Ronald Korn MD PhD (SMIL/Imaging Endpoints)

Abstract

We hypothesize that tumor heterogeneity or tissue complexity, as measured by quantitative texture analysis (QTA) on mammogram, is a marker of tumor aggressiveness in breast cancer patients. Tumor heterogeneity was assessed using QTA on digital mammograms of 64 patients with invasive ductal carcinoma (IDC). Tumor aggressiveness was assessed using patients' OncoType DX® Recurrence Score (RS), a proven genomic assay score that correlates with the 10-year rate of remote breast cancer recurrence and ranges from 1-100. Correlation between QTA parameters and RS is analyzed using statistical tools including student's t-test, Spearman rank correlation, linear regression, and logistic regression. Student's t-test shows that QTA parameter Mean has different average values for different RS risk groups in tumor tissues. Best-fit linear models show that QTA parameters at different filter levels, along with PR status if included, are good linear predictors of RS ($p < 0.0108$). Likewise, best-fit logistic models show that different QTA parameters at different filter levels, along with progesterone receptor (PR) status if included, are good predictors of high risk RS probability ($p < 0.0087$), using cutoffs of 30 or 25 for high risk RS. The results show that quantitative measurement of breast cancer tumor heterogeneity using QTA on digital mammograms can potentially allow a non-invasive and cost-effective way to quickly assess the likelihood of RS and high risk RS.

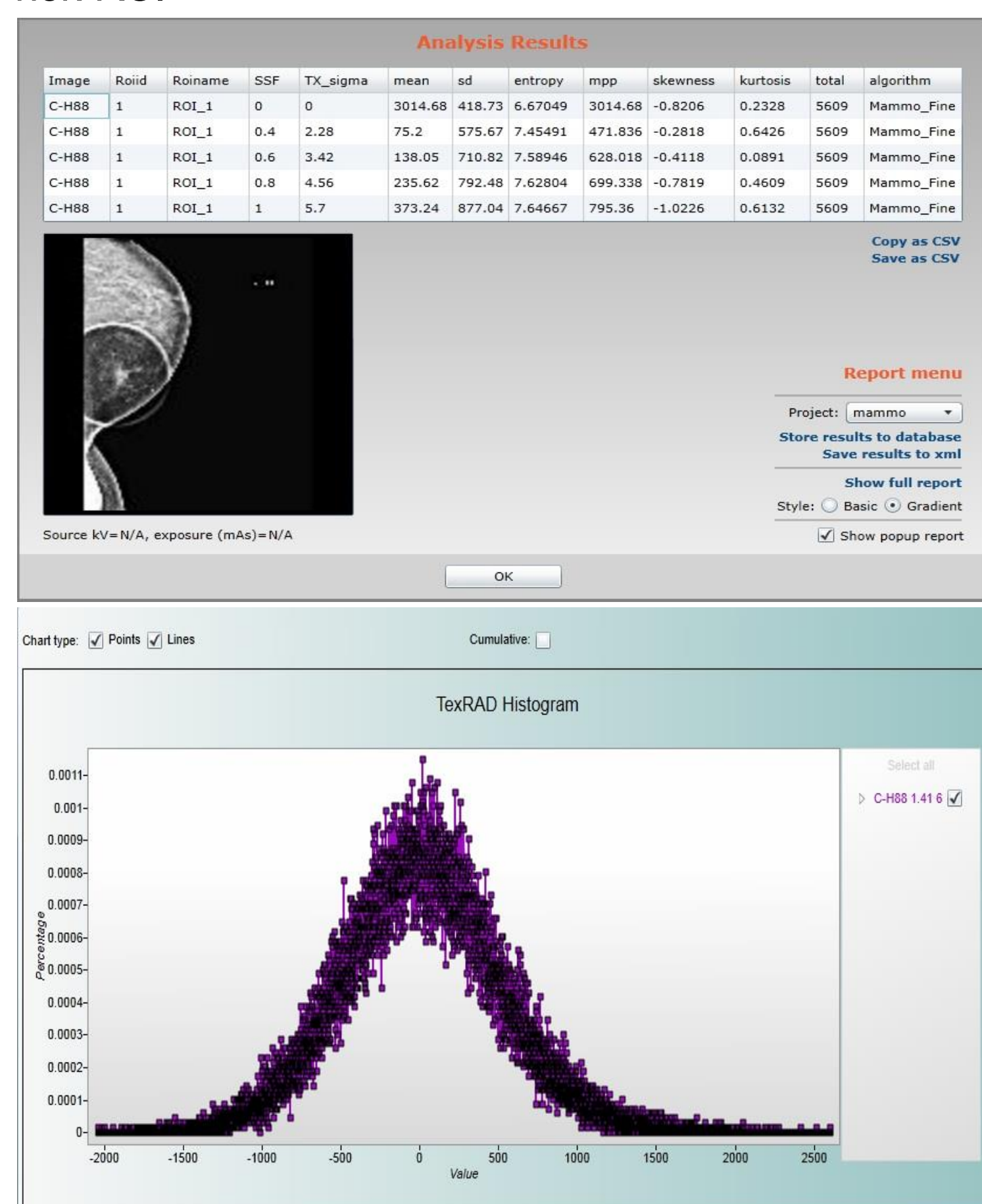


Figure 1: Sample QTA parameter output from TexRAD and the corresponding histogram output for a specific filter level. QTA parameters quantize the histogram output.

Introduction

Breast cancer is the second most common cancer among American women, second only to skin cancer. Complexity in treating breast cancer is due to heterogeneity of breast tumors, which is a result of Darwinian evolution of tumor. The heterogeneity is demonstrated by differences in genomic, proteomic, and metabolic expressions of tumor cells. Recognition of this heterogeneity has led to successful targeted therapy and management. Small pilot studies have demonstrated that texture analysis done on radiologic images of tumors correlates with tumor heterogeneity. Our goal is to examine the correlation between mammogram texture analysis and breast tumor aggressiveness.

Methods

This is a retrospective and exploratory study using archived digital mammography and OncoType DX® Recurrence Score (RS) of 64 women diagnosed with IDC. Average age is 61 (SD=11) and ranges from 36 to 83. Average RS is 20 (SD=11) and ranges from 4 to 65. 7 of them are PR- and only 1 of them is ER-. Quantitative texture analysis (QTA) is performed using a software program called TexRAD. TexRAD yields primary QTA parameters Mean, Standard Deviation, Entropy, Mean Positive Pixel Value, Skewness, and Kurtosis; from which 3 additional sets of QTA parameters are derived. QTA parameters quantize the distribution of gray level intensity of pixels in the user-defined regions of interest (i.e. tumor and normal tissue) at given filter levels. Student's t-test is used to test if QTA parameters differ among different RS risk groups. In addition, Spearman rank correlation, multiple linear regression, and multiple logistic regression are performed using QTA parameters, hormonal receptor status, and age as predictors, and RS or high risk RS ($RS > 30$ or $RS > 25$) as outcome.

Results: QTA vs RS

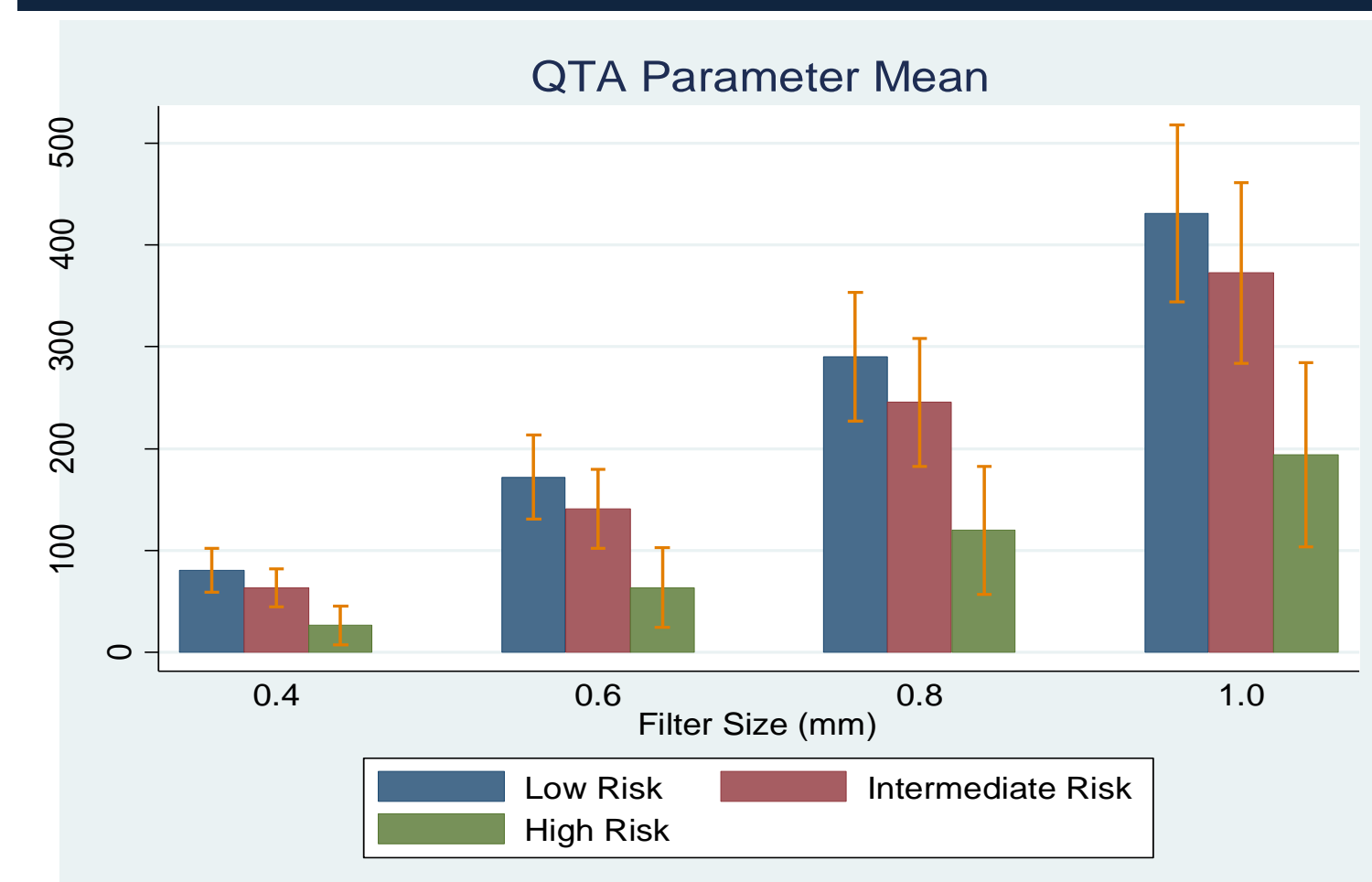


Figure 2: The mean and SD of QTA parameter Mean, for different RS risk groups (< 18 , $18-30$, and > 30) and filter levels (0.4 – 1.0mm) in tumor, demonstrating a negative correlation between QTA parameter Mean and high risk of breast tumor recurrence.

Using student's t-test, we observed that the average QTA parameter Mean for different RS risk groups differed most with statistical significance among three RS risk groups ($RS > 30$, $18 \leq RS \leq 30$, and $RS < 18$) ($t: 3.1-4.2$, $p = 0.0002-0.0057$). This phenomenon was not seen in normal tissue.

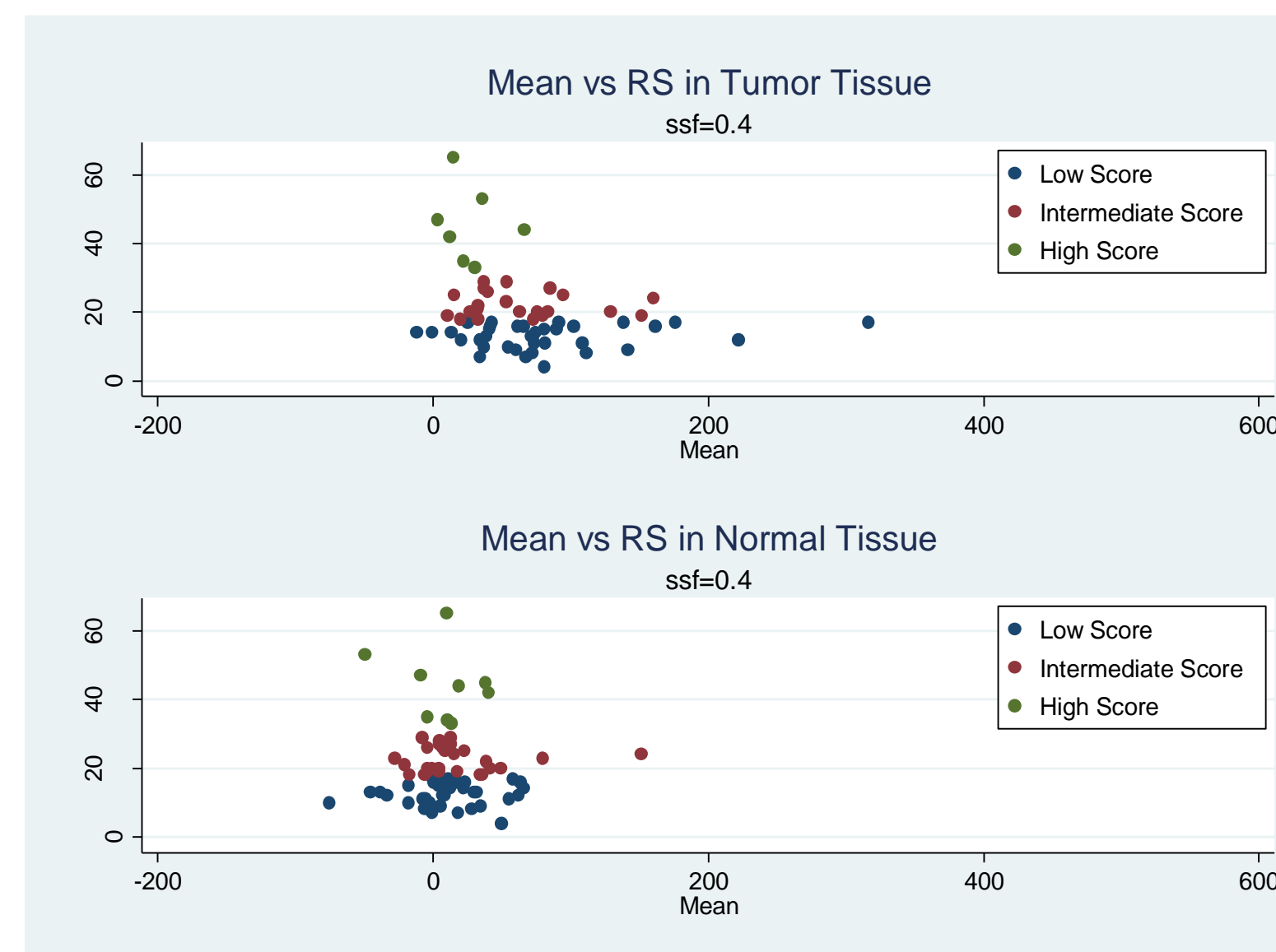


Figure 3: Scatter plot of RS vs. QTA parameter Mean in tumor vs. normal tissues, demonstrating tumor heterogeneity.

Results: Linear Regression to Predict RS

Multiple linear regression shows that QTA and PR status, and QTA alone can be used to predict $\log(RS)$. The best-fit linear model differs depending on whether PR status is included and/or unknown.

PR Status	SSF	Outcome	Predictors	Linear Model	n	F	p	R ²
Not Included	0.4	Log (RS)	SD	$\log(RS) = -0.0004177*(SD) + 1.500248$	65	6.89	0.0108	0.0870
Included	1	Log (RS)	PR, Skewness-Diff	$\log(RS) = -0.3759396*(PR) + 0.0785932*(Skewness-Diff) + 1.592273$	65	15.30	<	0.2988
PR+ Only	0.8	Log (RS)	Skewness-Diff	$\log(RS) = 0.3033226*(Skewness-Diff) + 2.83791$	58	9.36	0.0034	0.1320

Table 1: Best-fit QTA-based linear model in predicting $\log(RS)$, for 3 scenarios: a) PR included b) PR not included c) PR+ only. This shows that QTA may be used to predict RS. SSF is the filter level of the given best-fit model.

Results: Logistic Regression to Predict RS Risk Group

Multiple logistic regression shows that QTA and PR status, and QTA alone can be used to predict high risk RS ($RS > 30$ or $RS > 25$). The best-fit logistic model differs depending on whether PR status is included and/or unknown, and the RS cutoff value for high risk recurrence.

PR Status	Pr	Predictors	Logistic Model	SSF	n	chi ²	p
Not Included	$RS > 30$	Entropy-Diff	$p = \exp(x) / (1 + \exp(x))$ where $x = -5.905371*(Entropy-Diff) - 2.254296$	0	65	10.98	0.0009
Not Included	$RS > 25$	Mean-Total	$p = \exp(x) / (1 + \exp(x))$ where $x = -140.8101*(Mean-Total) - 0.2993403$	0.6	65	9.98	0.0016
Included	$RS > 30$	PR, SD-Diff	$\logit(p) = -3.548941*(PR) - 0.0092257*(SD-Diff) + 0.363196$	0	65	18.69	0.0001
Included	$RS > 25$	PR, Mean-Total	$\logit(p) = -4.321735*(PR) - 159.0879*(Mean-Total) + 3.430188$	0.6	65	25.56	<0.0001
PR+ Only	$RS > 30$	SD-Diff	$p = \exp(x) / (1 + \exp(x))$ where $x = -0.0124166*(SD-Diff) - 3.547383$	0	58	6.87	0.0087
PR+ Only	$RS > 25$	MPP-Diff, Skewness-Diff	$\logit(p) = -0.0095748*(MPP-Diff) + 6.487719*(Skewness-Diff) - 2.761748$	0.6	58	16.17	0.0003

Table 2: Best-fit QTA-based logistic model in predicting high risk RS group ($RS > 30$ or > 25), for 3 scenarios: a) PR included b) PR not included c) PR+ only. This shows that QTA may be used to predict high risk RS. Pr is the probability of having high risk RS (either $RS > 25$ or $RS > 30$).

Discussion and Conclusions

QTA Parameter Mean shows statistically significant correlation with RS risk groups in tumor tissue but not in normal tissue, confirming that QTA can be used to evaluate heterogeneity of tumor. Furthermore, multiple robust regression using QTA parameters and PR status, or QTA parameters alone are shown to be predictors for RS in linear model and high risk RS in logistic model. The findings confirm that QTA can be used to evaluate breast tumor heterogeneity and may be used to predict tumor aggressiveness. Our sample size is small. Future directions include increasing sample size, performing QTA on whole breast tissue, studying correlation between QTA and microcalcification lesions, as well as using QTA to predict hormonal receptor (PR, ER) status.

Acknowledgements

I wish to thank my mentor Dr. Ronald Korn and his collaborator Dr. Jasjit Sachdev. I would also like to thank biostatisticians Dr. Chengcheng Hu and Paul Kang for reviewing the statistical analysis in this study.