Impact of Tumor Volume Transformations on Prediction of Breast Cancer Recurrence

Kevin Zheng | John Kornak | Wen Li | Natsuko Onishi | Nola Hylton

May 2024

1 Introduction

Breast cancer is the second most common cancer in the world and remains one of the leading causes of cancer-related death. Despite advances in treatment, prediction of patient outcomes, particularly risk of recurrence, remains a challenge in breast cancer management. Neoadjuvant chemotherapy (NAC), which refers to chemotherapy treatment administered before surgery, has become an important staple in treatment methodologies [Wang et al. 2017]. NAC has started to replace the previously standard surgery-first treatment approach. It offers key advantages over postsurgical chemotherapy, such as the ability to monitor treatment response and tailor medical interventions accordingly [Li et al. 2022]. This approach also facilitates a more powerful acquisition of research data, as treatment responses can be monitored over time. Thus, the I-SPY 2 breast cancer trial, a study aiming to evaluate the efficacy of experimental drug treatments in NAC, has provided important information through its systematic tracking of patients during their intervention period.

An important aspect of this trial is the use of dynamic contrast-enhanced (DCE) breast magnetic resonance imaging (MRI) to monitor changes in tumor response. Among the various metrics extracted from these MRI scans, functional tumor volume (FTV), a quantitative measure of tumor burden derived from DCE breast MRI at four

time points, has been shown to be a promising biomarker for predicting treatment response [Hylton et. Al 1999]. In this analysis, FTV3 reflects FTV obtained from the fourth MRI taken at the end of NAC. Previous studies, including research performed on the earlier I-SPY 1 trial, found that FTV predicts a pathologic complete response (pCR) and overall survival outcomes [Hylton et al. 2016]. However, potential transformations of FTV values might be more predictive in the context of survival analysis, especially when considered in conjunction with the breast cancer subtype.

This study aims to test this hypothesis by investigating how transformed versions of FTV3 (log and cube root transformations) compared to the untransformed version of FTV3 in predicting breast cancer recurrence through Cox proportional hazards models and time-dependent receiver operator characteristic (ROC) curves. Log and cube root transformations are commonly used in the normalization of continuous data [Osborne et al. 2002], serving to potentially normalize right skewness in datasets. The transformation of the cube root ($\sqrt[3]{x}$) was of particular interest because tumor volume can be interpreted as the product of three-dimension components [Bektaş et al. 2021].

To protect against overfitting, we applied k-fold cross-validation (CV) to evaluate our Cox models, with c-indices as our validation metric. These models will also incorporate the breast cancer receptor status based on hormone receptor (HR) and Human Epidermal Growth Factor Receptor 2 (HER2), classifying patients into the following four subtypes: "HR-/HER2-," "HR+/HER2-," "HR-/HER2+," and "HR+/HER2+." These two receptor statuses are critical biomarkers in determining optimal treatment strategy of breast cancer [Korde et. al 2021].

Once the “best” (transformed) version of FTV3 is determined, we evaluate the Cox model fits based on the full data and subtype-specific models fitted to data subsets for each cancer subtype.

In summary, the goals of this study are to (1) determine whether transformations improve survival prediction with cross-validated Cox proportional hazards models and time-dependent ROC curves; (2) compare hazard ratios with each of the four statuses as the reference group using a Cox model with the best performing (transformed) version of FTV3; (3) compare Cox models for each breast cancer status in subgroup analyses with the best (transformed) version of FTV3.

2 Methods

2.1 I-SPY 2 Dataset

The dataset for this project is derived from the first subset of publicly available imaging data from the I-SPY 2 breast cancer trial, known as the "I-SPY2 Imaging Cohort 1" archived in the Cancer Imaging Archive (TCIA) [Li et al. 2022]. This subset comprises 985 participants, all women with locally advanced breast cancer (aged 18 and above) enrolled in the I-SPY 2 trial from 2010 to 2016. The cohort has a median age of 49 years (range 23 to 77 years). Each subject was adaptively randomized to one of nine experimental drug treatment arms or the control arm. MRI scans were taken of each individual at four different time points during their neoadjuvant chemotherapy treatment periods: pre-treatment, post-cycle 3 of drug administration, post-cycle 12 (final cycle), and pre-surgery after the completion of all treatments [Li et al. 2022]. Functional tumor volumes (FTV3) used in this analysis, are derived from the pre-surgery MRI scans.

For the final analysis, a total of 899 participants were included based on the availability of long-term survival data. FTV3 data was missing for 79 participants, and the missingness appeared to be randomly distributed without identifiable patterns. To address missingness, we performed imputation using predictive mean matching with the MICE package in R.

The cohort was further categorized by subtype based on the HR and HER2 status, which were distributed as follows: 79 individuals were HR-/HER2+, 323 individuals were HR-/HER2-, 357 individuals were HR+/HER2-, and 140 individuals were HR+/HER2+. Additional details describing the cohort are presented in Table 1.

Table 1: Demographics and Treatment Arms Overview

	Overall (N=899)
Age Mean (SD)	48.8 (10.5)
Median [Min, Max]	49.0 [23.0, 77.0]
Race	
Asian	62 (6.9%)
Black or African American	103 (11.5%)
Mixed Heritage or Indigenous Peoples	16 (1.8%)
White	718 (79.9%)
Breast Cancer Subtype	
HR-/HER2-	323 (35.9%)
HR-/HER2+	79 (8.8%)
HR+/HER2-	357 (39.7%)
HR+/HER2+	140 (15.6%)
Menopausal Status	
Premenopausal	440 (48.9%)
Perimenopausal (around menopause)	34 (3.8%)
Postmenopausal	275 (30.6%)
Above categories not applicable AND Age < 50	44 (4.9%)
Above categories not applicable AND Age > 50	74 (8.2%)
NA	32 (3.6%)
Treatment Arm	
Paclitaxel	169 (18.8%)
Paclitaxel + ABT 888 + Carboplatin	63 (7.0%)
Paclitaxel + AMG 386	104 (11.6%)
Paclitaxel + AMG 386 + Trastuzumab	17 (1.9%)
Paclitaxel + Ganetespib	88 (9.8%)

Paclitaxel + Ganitumab	97 (10.8%)
Paclitaxel + MK-2206	50 (5.6%)
Paclitaxel + MK-2206 + Trastuzumab	28 (3.1%)
Paclitaxel + Neratinib	104 (11.6%)
Paclitaxel + Pembrolizumab	64 (7.1%)
Paclitaxel + Pertuzumab + Trastuzumab	38 (4.2%)
Paclitaxel + Trastuzumab	30 (3.3%)
T-DM1 + Pertuzumab	47 (5.2%)

2.2 Statistical Analysis

The statistical modeling focused on survival outcomes of recurrence of breast cancer. The main predictor of interest is FTV3, with transformed variations also considered as alternative predictors such as cube root FTV3, and log FTV3. The combined breast cancer status indicates the status for a set of biomarkers (e.g., HR and HER2 status), which is included as a covariate.

Five-fold cross-validation was performed on three Cox proportional hazards models to assess the predictive accuracy of the different transformations of FTV3. The concordance index (C-index) was used as the evaluation metric to select the best model. The models considered were:

Model 1: Cox model with FTV3 and combined status as predictors.Model 2: Cox model with cube root FTV3 and combined status as predictors.

Model 3: Cox model with log FTV3 and combined status as predictors.

The model chosen for downstream analysis was based on a combination of the c-indices performance and interpretability.

For each of the three models, time-dependent ROC curves were constructed for survival outcomes at 1 year, 2 years, and 5 years using the PROC PHREG procedure in

SAS. The performance of each model was evaluated by comparing the area under the curve (AUC) at each time point (1 year, 2 years, and 5 years).

For the best-performing model based on cross-validated ROC curve performance, a Cox proportional hazards model was used to estimate hazard ratios for FTV3 and combined status. The estimates, 95% confidence intervals, and p-values were calculated. Additionally, we conducted all six pairwise comparisons between the following four combined status groups:

“HR-/HER2-"

“HR+/HER2-"

“HR-/HER2+"

“HR+/HER2+"

Given the reduced sample sizes and a priori expected correlations among results, we did not correct for multiple comparisons. Any between-subgroup comparison results should therefore be taken as exploratory, requiring independent confirmation.

To examine the potential interaction between FTV3 and combined status, a Cox model incorporating an interaction term between FTV3 (in its best transformation) and combined status was fit. The significance of the interaction term was evaluated with the (partial) likelihood ratio test by comparing the interaction model to a main-effects-only model (i.e., without the interaction term).

Given the possibility that the effect of FTV3 on survival may differ by combined status subgroup, subgroup analyses were performed. The data were stratified into four subsets based on combined status, and separate Cox models were fit for each subset, using FTV3 (in its best transformation) as the sole predictor.

All analysis was performed in R version 4.4.0 and SAS version 9.4.

3 Results

As shown in Table 2, the cross-validated c-indices for the two transformed versions of FTV3 performed similarly (log FTV3 c-index: 0.698, cube root FTV3 c-index: 0.694)

when utilizing the imputed dataset. Both transformations performed better than untransformed FTV3, which yielded a c-index of 0.668.

A sensitivity analysis excluding missing values for FTV3 (n = 820) yielded similar results, with log FTV3 c-index = 0.7057 and cube root FTV3 c-index = 0.7023. Both transformed models again performed better than the untransformed FTV3, with c-index = 0.6781. Despite the small increase in performance of log FTV3 over cube root in these tests, cube root was selected for subsequent analysis due to having greater interpretability as representing a volume.

Table 2: 5 Fold CV Results

Variable (with combined status)	C-index (imputed data)	Rank Imputed	C-index (drop NA)	Rank Drop NA
log FTV3	0.6982	1	0.7057	1
cube root FTV3	0.6935	2	0.7023	2
non transformed FTV3	0.6681	3	0.6781	3

Time-dependent ROC curves for the three forms of FTV3, combined with cancer status, are shown in Fig. 1. Both the log and cube root transformations of FTV3 demonstrated higher AUCs compared to the untransformed FTV3 for 1-year, 2-year, and 5-year survival. For instance, at 1-year survival, the AUCs for log FTV3, cube root FTV3, and raw FTV3 were 0.8264, 0.8217, and 0.807, respectively. The AUC values for 2-year and 5-year survival are also presented in Fig. 1, where a clear decreasing trend in AUC values is observed as survival time increases. Fig. 2 presents time-dependent ROC curves for the cube root FTV3 model, adjusted for breast cancer status, at 1-year, 2-year, and 5-year intervals. These curves illustrate a similar decreasing trend, with AUCs of 0.82,

0.76, and 0.65 at 1-year, 2-year, and 5-year survival, respectively.

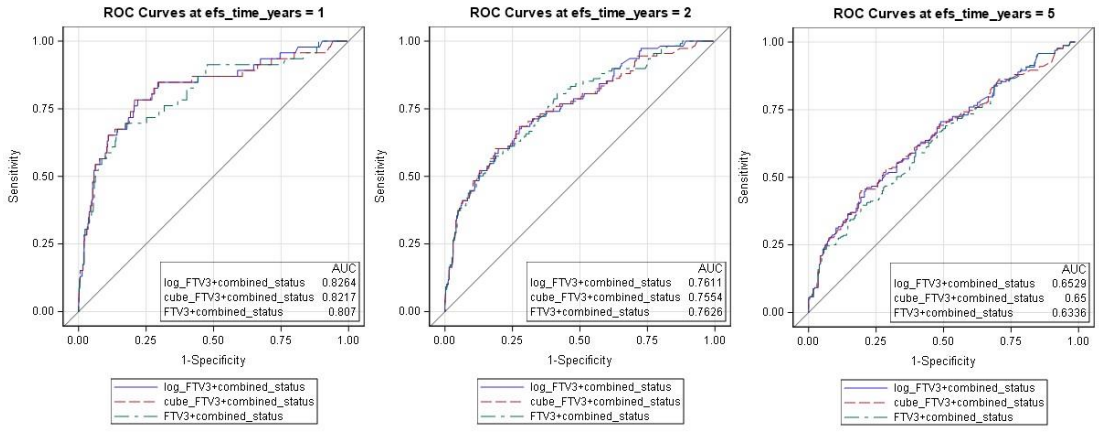

Figure 1: Overlapped ROC Curves (With Status) by Year

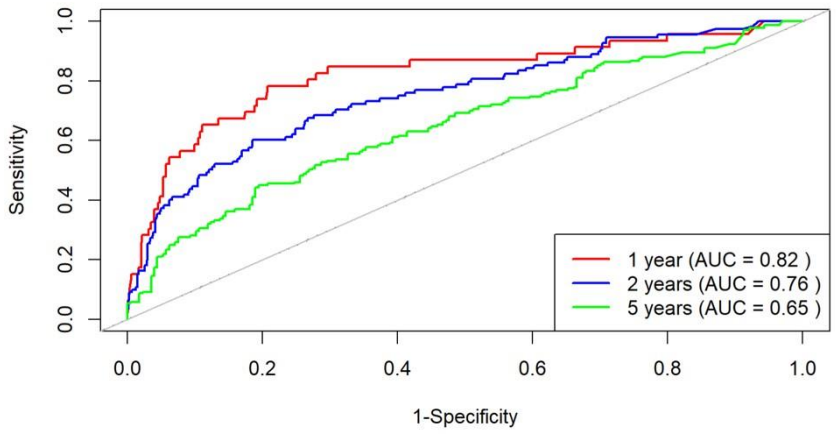

Figure 2: Cube Root FTV3 (with status) 1,2,5 year Overlapped ROC Curve

Table 3 shows the table of results from the pairwise comparisons, which are reported with hazard ratios, p values, and 95% confidence intervals. The estimated hazard ratio for cube-root FTV3 was 1.957 (95% CI: 1.731-2.212, $p < 0.0001$) when adjusted for breast cancer status. Among the different subgroups examined, only HR-/HER2- vs HR+/HER2+ reached a statistically significant difference (hazard ratio 0.614, 95%CI: 0.378-0.999, $p = 0.049$).

Table 3: Pairwise Comparisons

Comparison	Hazard Ratio	p-value	Lower .95	Upper .95
HR-/HER2- vs HR+/HER2-	0.854	0.349	0.614	1.188
HR-/HER2- vs HR-/HER2+	0.881	0.646	0.514	1.512
HR-/HER2- vs HR+/HER2+	0.614	0.049	0.378	0.999
HR+/HER2- vs HR-/HER2+	1.032	0.910	0.598	1.781
HR+/HER2- vs HR+/HER2+	0.719	0.188	0.441	1.174
HR-/HER2+ vs HR+/HER2+	0.697	0.278	0.363	1.338
cube_FTV3	1.957	<2e-16	1.731	2.212

When we fit the Cox model with cube-root FTV3 and breast cancer status interaction term, the p value for the interaction term (ANOVA) was 0.83. Although this does not rule out the possibility of an interaction effect, we adopt the more parsimonious model without it.

In the single predictor Cox model analysis using subsetted data by cancer subtype (Table 4), cube root FTV3 is a statistically significant predictor in all four subtypes.

Table 4: Cox Models by Subgroup

Subgroup	Hazard Ratio	p-value	Lower .95	Upper .95	C-Index
HR-/HER2-	1.819	1.23E-11	1.530	2.163	0.668
HR+/HER2-	2.107	2.49E-09	1.649	2.692	0.698
HR-/HER2+	1.823	0.000257	1.321	2.516	0.712
HR+/HER2+	2.170	5.60E-05	1.489	3.163	0.770

4 Discussion

The main findings of this study are that (1) both log-transformed FTV3 and cube-root FTV3 improve survival prediction compared to untransformed FTV3 when examined using cross-validated Cox proportional hazards models. (2) The two transformed forms of FTV3 outperformed untransformed FTV3 in predicting survival in the Cox models built using imputed data. The results were similar to those built using the data with missing data excluded, suggesting that imputation did not influence the relative predictability of the two transformed forms compared to the untransformed FTV3. (3) The log FTV3 model performed slightly better than the cube root FTV3 model.

However, the differences between the log FTV3 and cube-root FTV3 models were minimal, suggesting that both transformations performed similarly. Cube-root FTV3 was chosen due to its greater interpretability as representing a volumetric measure converted to an approximate length. (4) In the time-dependent ROC curve analyses, we observed that the log FTV3 and cube-root FTV3 generated higher AUC than untransformed FTV3 at 1 year, 2 year, and 5 year survival. The AUC for all three forms of FTV3 showed a decreasing trend as the survival time increased from 1 to 5 years. This would generally be expected since current conditions are likely to have reduced predictive value further into the future. (5) We did not see evidence for a clear relationship between cube root FTV3 and breast cancer. Although such an interaction cannot be ruled out and should continue to be investigated, we proceeded on the basis of the additive effect model being adequate. (6) In the subset analyses, we found that cube-root FTV3 is a statistically significant predictor in all 4 breast cancer subgroups.

While our study shows promising results for cube-root transformations of FTV3 in improving survival prediction, it is important to consider potential limitations. One limitation is that while the cube-root transformation displayed similar predictive value to the log transformation, and improved predictive power over untransformed FTV3, this should be further evaluated in independent cohorts. The generalizability of these findings may be limited by participant characteristics, such as age, treatment regimens, menopausal status, or other factors that were not accounted for in this study. Another potential limitation lies in the use of imputed data. While the imputation method used in this study was well validated, the choice of imputation strategy and the potential impact of unmeasured biases or patterns in the missing data should be considered. Future work could be conducted to examine the sensitivity of the results to different imputation methods. Additionally, studies may be done to investigate whether certain subgroups (such as cancer subtype or treatment type) are more sensitive to missing data.

In conclusion, our results provide strong evidence that cube root FTV3 is a better predictor of breast cancer survival than untransformed FTV3.

5 References

1. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning with Applications in R: By Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani,. Vol 6. Springer; 2023.

2. Marshall A, Altman DG, Holder RL. Comparison of imputation methodsfor handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. BMC Med Res Methodol. 2010;10(1):112. doi:10.1186/1471-2288-10-112

3. Blanche P, Dartigues J, Jacqmin-Gadda H. Estimating and comparingtime-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Statistics in Medicine. 2013;32(30):5381-5397. doi:10.1002/sim.5958

4. Van Buuren S. Flexible Imputation of Missing Data, Second Edition. 2nded. Chapman and Hall/CRC; 2018. doi:10.1201/9780429492259

5. Li, W., Newitt, D. C., Gibbs, J., Wilmes, L. J., Jones, E. F., Arasu, V. A., Strand, F., Onishi, N., Nguyen, A. A.-T., Kornak, J., Joe, B. N., Price, E. R., Ojeda-Fournier, H., Eghtedari, M., Zamora, K. W., Woodard, S. A., Umphrey, H., Bernreuter, W., Nelson, M., … Hylton, N. M. (2022). I-SPY 2 Breast Dynamic Contrast Enhanced MRI Trial (ISPY2) (Version 1) [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/TCIA.D8Z0-9T85

6. Vaz-Luis I, Ottesen RA, Hughes ME, et al. Impact of hormone receptor status on patterns of recurrence and clinical outcomes among patients with human epidermal growth factor-2-positive breast cancer in the National Comprehensive Cancer Network: a prospective cohort study. Breast Cancer Res. 2012;14(5):R129. doi:10.1186/bcr3324

7. Abd ElHafeez S, D'Arrigo G, Leonardis D, Fusaro M, Tripepi G, Roumeliotis S. Methods to Analyze Time-to-Event Data: The Cox Regression Analysis. Georgakilas A, ed. Oxidative Medicine and Cellular Longevity. 2021;2021:1-6. doi:10.1155/2021/1302811

8. Buuren SV, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Soft. 2011;45(3). doi:10.18637/jss.v045.i03

9. Kleinke K. Multiple Imputation Under Violated Distributional Assumptions: A Systematic Evaluation of the Assumed Robustness of Predictive Mean Matching. Journal of Educational and Behavioral Statistics. 2017;42(4):371-404. doi:10.3102/1076998616687084

10. Wang M, Hou L, Chen M, et al. Neoadjuvant Chemotherapy Creates Surgery Opportunities For Inoperable Locally Advanced Breast Cancer. Sci Rep. 2017;7(1):44673. doi:10.1038/srep44673

11. Hylton NM, Gatsonis CA, Rosen MA, et al. Neoadjuvant Chemotherapy for Breast Cancer: Functional Tumor Volume by MR Imaging Predicts Recurrence-free Survival—Results from the ACRIN 6657/CALGB 150007 I-SPY 1 TRIAL. Radiology. 2016;279(1):44-55. doi:10.1148/radiol.2015150013

12. Osborne J. Notes on the use of data transformations. doi:10.7275/4VNG5608

13. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statisticsfor evaluating overall adequacy of risk prediction procedures with censored survival data. Statistics in Medicine. 2011;30(10):1105-1117. doi:10.1002/sim.4154

14. Hartman N, Kim S, He K, Kalbfleisch JD. Pitfalls of the concordanceindex for survival outcomes. Statistics in Medicine. 2023;42(13):21792190. doi:10.1002/sim.9717

15. Shi Y, Olsson LT, Hoadley KA, et al. Predicting early breast cancer recurrence from histopathological images in the Carolina Breast Cancer Study. npj Breast Cancer. 2023;9(1):92. doi:10.1038/s41523-023-00597-0

16. Bektaş AB, Gönen M. PrognosiT: Pathway/gene set-based tumour volumeprediction using multiple kernel learning. BMC Bioinformatics. 2021;22(1):537. doi:10.1186/s12859-021-04460-6

17. Al-Issa Y, Mohammad Alqudah A, Alquran H, Al Issa A. PulmonaryDiseases Decision Support System Using Deep Learning Approach. Computers, Materials & Continua. 2022;73(1):311-326. doi:10.32604/cmc.2022.025750

18. Harrell , FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer International Publishing; 2015. doi:10.1007/978-3-319-19425-7

19. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B: Statistical Methodology. 1972;34(2):187-202. doi:10.1111/j.2517-6161.1972.tb00899.x

20. Understanding Cancer Prognosis. Accessed July 2, 2024. https://www.cancer.gov/about-cancer/diagnosis-staging/prognosis

21. Hylton NM. Vascularity assessment of breast lesions with gadolinium-enhanced MR imaging. *Magn Reson Imaging Clin N Am*. 1999;7(2):411-420, x.

22. Harris MA, Savas P, Virassamy B, et al. Towards targeting the breast cancer immune microenvironment. *Nat Rev Cancer*. 2024;24(8):554-577. doi:10.1038/s41568-024-00714-6

23. Korde LA, Somerfield MR, Carey LA, et al. Neoadjuvant Chemotherapy, Endocrine Therapy, and Targeted Therapy for Breast Cancer: ASCO Guideline. JCO. 2021;39(13):1485-1505. doi:10.1200/JCO.20.03399