

STAT 306 Interim Proposal

Nicholas Tam, Ivy Cheung, Kaichi Nakajima, Kevin Liu (Group 10)

Variables

- Data obtained from Kaggle dataset, which in turn was obtained from Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) database through CBioPortal for Cancer Genomics.
- Variables from columns 32 to 693 consist of genetic attributes containing m-RNA levels z-score for 331 genes, and mutation for 175 genes; they have been omitted due to being difficult to interpret.

Variable	Definition	Unit	Categories
age_at_diagnosis	Age of the patient at diagnosis time	Years	
type_of_breast_surgery	Breast cancer surgery type	N/A	- "MASTECTOMY", "BREAST CONSERVING"
cancer_type_detailed	Detailed breast cancer types	N/A	"Breast Invasive Ductal Carcinoma", "Breast Mixed Ductal and Lobular Carcinoma", "Breast Invasive Lobular Carcinoma", "Breast Invasive Mixed Mucinous Carcinoma", "Metaplastic Breast Cancer"
chemotherapy	Boolean on whether or not patient had chemotherapy as a treatment	Boolean	
hormone_therapy	Whether or not the patient had hormonal therapy as a treatment	Boolean	
overall_survival_months	Duration from the time of the intervention to death	Months	
overall_survival	Whether the patient is alive or dead by the time of recording	Boolean	
radio_therapy	Whether or not the patient had radiotherapy as a treatment	Boolean	
tumor_size	Tumor size measured by imaging techniques	mm	
tumor_stage	Stage of cancer based on involvement of surrounding structures, lymph nodes and distant spread	N/A	- 0, 1, 2, 3, 4, N/A

Research Question

Breast invasive ductal carcinoma (IDC) is the most common type of breast cancer, with about 80% of all forms of breast cancer being IDC, according to the American Cancer Society (DePolo, 2024). There are numerous nonsurgical treatments of IDC, such as radiotherapy, chemotherapy, and hormone therapy (Wright, 2023), but it is unclear how combinations of treatments can interact in a model to predict a patient's survival until death.

Our project question is: While factoring in age, tumor size and tumor stage, how do radiotherapy, chemotherapy, and hormone therapy influence the length of time a patient with IDC will survive until death?

Due to the distribution of categories and for ease of computation, we will be filtering the dataset for deceased IDC patients only, as there are inconsistent inputs in the "overall_survival" column for living patients.

Overview

Each of us will create at least one visualisation with a corresponding analysis. One of us will create the introduction, another will fit the linear model of interest based on previous visualisations, and the remaining two members will cooperate to create the conclusion.

Source

- *Breast cancer gene expression profiles (METABRIC)*. Kaggle. (2016, May 10). <https://www.kaggle.com/datasets/raghadalharbi/breast-cancer-gene-expression-profiles-metabric>
- CBioPortal for Cancer Genomics. (n.d.). https://www.cbioportal.org/study/summary?id=brca_metabric
- DePolo, J. (2024, October 2). *Invasive ductal carcinoma (IDC)*. Breastcancer.org - Breast Cancer Information and Support. <https://www.breastcancer.org/types/invasive-ductal-carcinoma>
- Wright, P. (2023, March 21). *Invasive ductal carcinoma (IDC)*. Johns Hopkins Medicine. <https://www.hopkinsmedicine.org/health/conditions-and-diseases/breast-cancer/invasive-ductal-carcinoma-idc#:~:text=Radiation%20therapy%20might%20be%20part,lymph%20nodes%2C%E2%80%9D%20Wright%20says.>