The role of Age, Pathologic Stage, and Vital Status for Female Patients when Developing Breast Cancer

Mengdi Chai

06 December 2022

Introduction

Except for skin cancers, breast cancer is the most prevalent cancer in women in the US. 30% (or 1 in 3) of all new cases of female cancer each year are caused by it. Women in their middle years and older tend to develop breast cancer. At the time of breast cancer diagnosis, the average age is 62. As a result, 62 years of age or younger is the average age at which breast cancer in women is discovered. Women under 45 who have been diagnosed with breast cancer are extremely rare. In general, a woman's lifetime chance of acquiring breast cancer in the United States is around 13%. Because of the high prevalence and mortality rate of lung cancer, it is essential to discover the pathway and mechanism of breast cancer in female patients. The development of breast cancer is divided into several stages: stage I, stage II, stage III, stage IV, and stage X.

Breast cancer in stage I is an early stage and refers to a tiny, localized tumor that has only spread to the breast tissue or may have spread to nearby lymph nodes. In stage, the breat cancer can further be divided into IA stage and IB stage. A malignancy that is at stage IA is less than 2 centimeters (cm) in size and has not spread outside the breast, whereas Stage IB means that small areas of breast cancer cells are found in the lymph nodes close to the breast. Breast cancer that has reached stage II has spread to adjacent lymph nodes, the breast, or both. It is breast cancer in its early stages, and Stage II breast cancer has two groups: stage IIA and IIB. The classification of each group is depends on the size of tumour.

Stage III indicates that the cancer has migrated from the breast to nearby lymph nodes, the breast's surface, or the chest wall. Additionally, it is known as locally advanced breast cancer.Similarly, stage III also has different groups: stage IIIA, stage IIIB, and stage IIIC. Stage IIIA is quite complex since it depends on the size of tumour and the exist of breast cancer cells in the lymph nodes. Different from stage IIIA, Stage IIIB means the tumour has spread to the skin of the breast or the chest wall. In stage IIIC, there may or may not be a tumor, depending on its size. However, there is cancer in the skin of the breast that has gone to the chest wall and is causing edema or ulcers. Stage IV cancers have spread (metastasized) beyond the breast and nearby lymph nodes to other parts of the body.

The Cancer Genome Atlas Program (TCGA) is a cancer genomics program that collects over 20,000 cancer tissue samples and matches them with normal samples spanning over 33 cancer types (National Institute of Health). This publicly available database is used for multi-omic data analysis, which integrates data sets from many omic groupings.

In this study, I explored the survival rates of breast cancer in different age group at different pathologic stage to identify whether age and pathologic stage does play a role in patient outcomes. I hypothesized that age and pathologic stage will affect the female patient outcomes then they develop breast cancer.

Method

The analysis was mainly conducted in R using breast cancer clinical data. The data was sourced from TCGA with the TCGAbiolinks library using accession code "BRCA." In addition, another clinic

data was downloaded through TCGA website. Two data frames were merged together to access to "age_at_initial_pathologic_diagnosis", "ajcc_pathologic_stage", and "vital_status" columns.

In the dataset, there were a total of 2268 female patients. "age_category" column was created to divide female patients into young and old groups. According to the report of breast cancer, at the time of breast cancer diagnosis, the average age is 62. Therefore, female patients equal to or above 62 years old were categorized as "old", whereas female patients under 62 years old were categorized as "young". Missing data in "age_at_initial_pathologic_diagnosis" was excluded. Also, stage '– and stage X were excluded since they have no meaning.

The visualizations and analyses were created using an assortment of libraries within R. In R, TCGABioLinks was used to access the data. Survival and Survminer were used to create Kaplan-Meier Curves, ggplot2 was used to create bar plot, and plotly was used to create interactive plots.

Results

The bar plot in Figure 1 depicts the distribution of pathologic stage bewteen old female patients and young female patients. In each category, most female patients are in stage IIA or stage IIB, whereas stage II and stage III have the least number of people.

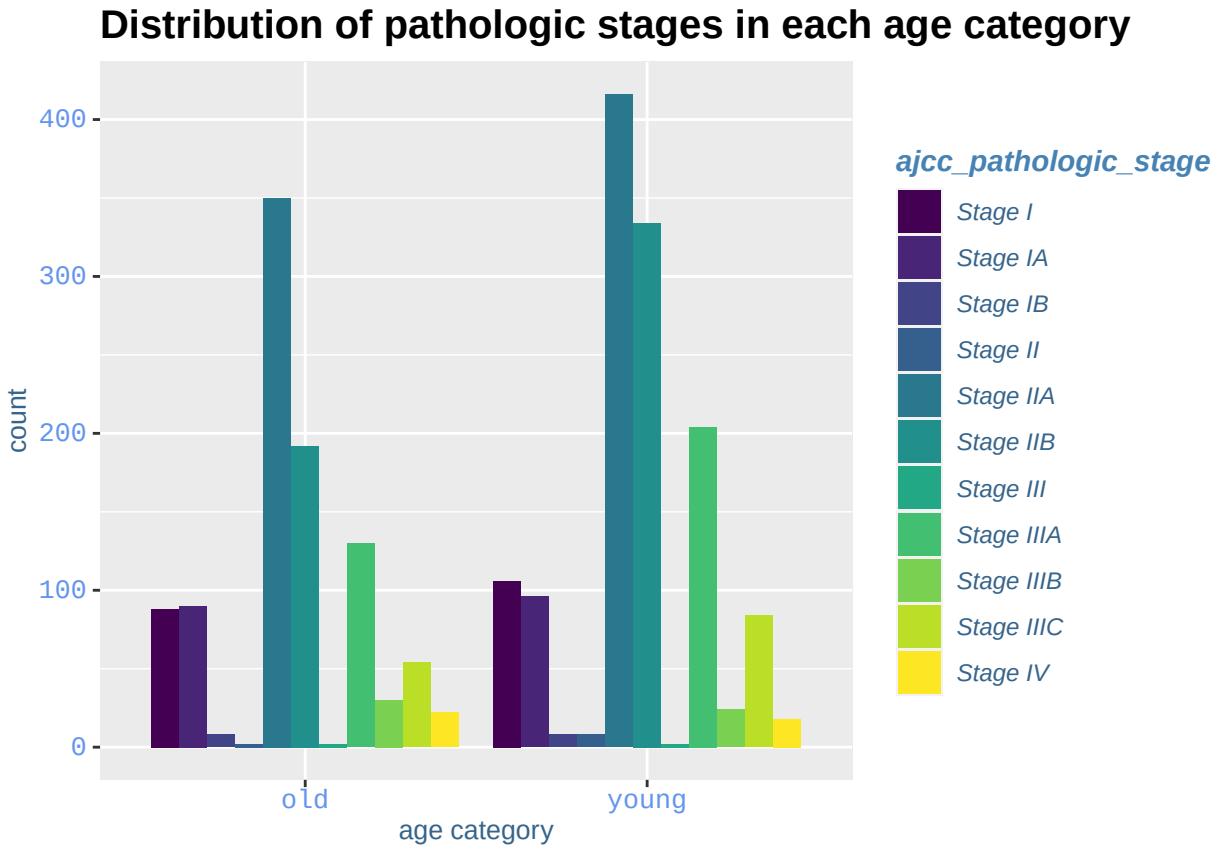

Figure 1. Bar plot for the distribution of pathologic stage bewteen old female patients and young female patients. There are 88 young female patients and 106 old female patients in stage I. There are 90 young female patients and 96 old female patients in stage IA. There are 8 young female patients and 8 old female patients in stage IB. There are 2 young female patients and 8 old female patients in stage II. There are 350 young female patients and 416 old female patients in stage IIA. There are 192 young female patients and 334 old female patients in stage IIB. There are 2 young female patients and 2 old female patients in stage III. There are 130 young female patients and 204 old female patients in stage IIIA. There are 30 young female

patients and 24 old female patients in stage IIIB. There are 54 young female patients and 84 old female patients in stage IIIC. There are 22 young female patients and 18 old female patients in stage IV.

Then, the distribution of vital status in each pathologic stage is examined. From Figure 2, all female patients in stage IB and stage II are alive, whereas the proportion of female patients whose vital status is dead at stage III, and stage IV is high.

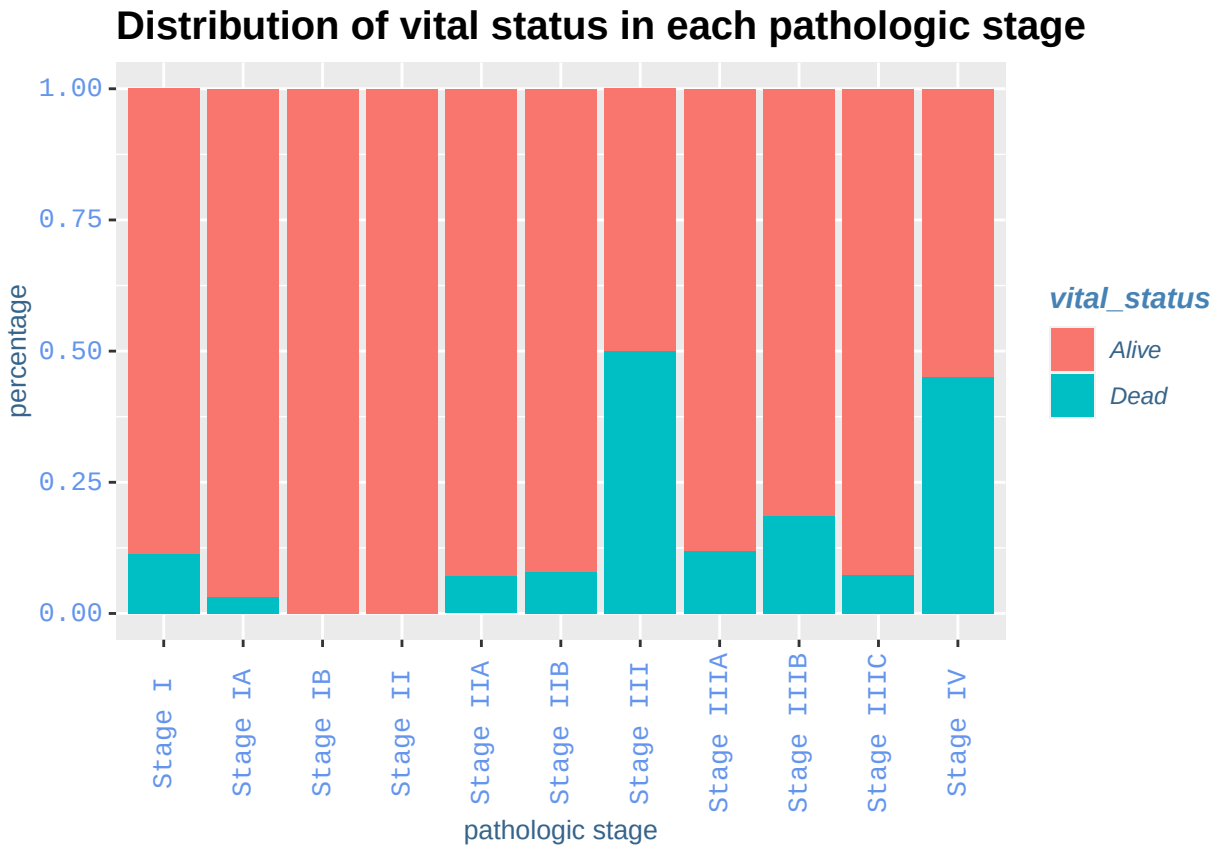

Figure 2. Bar plot for the distribution of vital status in each pathologic stage. The percentage of "alive" female patients is 89% and the percentage of "dead" female patients is 11% at stage I. The percentage of "alive" female patients is 97% and the percentage of "dead" female patients is 3% at stage IA. The percentage of "alive" female patients is 100% for both stage IB and stage II. The percentage of "alive" female patients is 93% and the percentage of "dead" female patients is 7% at stage IIA. The percentage of "alive" female patients is 92% and the percentage of "dead" female patients is 8% at stage IIB. The percentage of "alive" female patients is 50% and the percentage of "dead" female patients is 50% at stage III. The percentage of "alive" female patients is 88% and the percentage of "dead" female patients is 12% at stage IIIA. The percentage of "alive" female patients is 81% and the percentage of "dead" female patients is 19% at stage IIIB. The percentage of "alive" female patients is 93% and the percentage of "dead" female patients is 7% at stage IIIC. The percentage of "alive" female patients is 55% and the percentage of "dead" female patients is 45% at stage IV.

The scatter plot is draw to show the relationship between age, pathologic stage, and vital status. From figure 3, no relationship between aging and female patient's vital status. But for the pathologic stage, especially in stage IV, the proportion of patients dying is increasing. If putting the age factor in stage IV, the proportion of old patients whose vital status is dead is the highest.

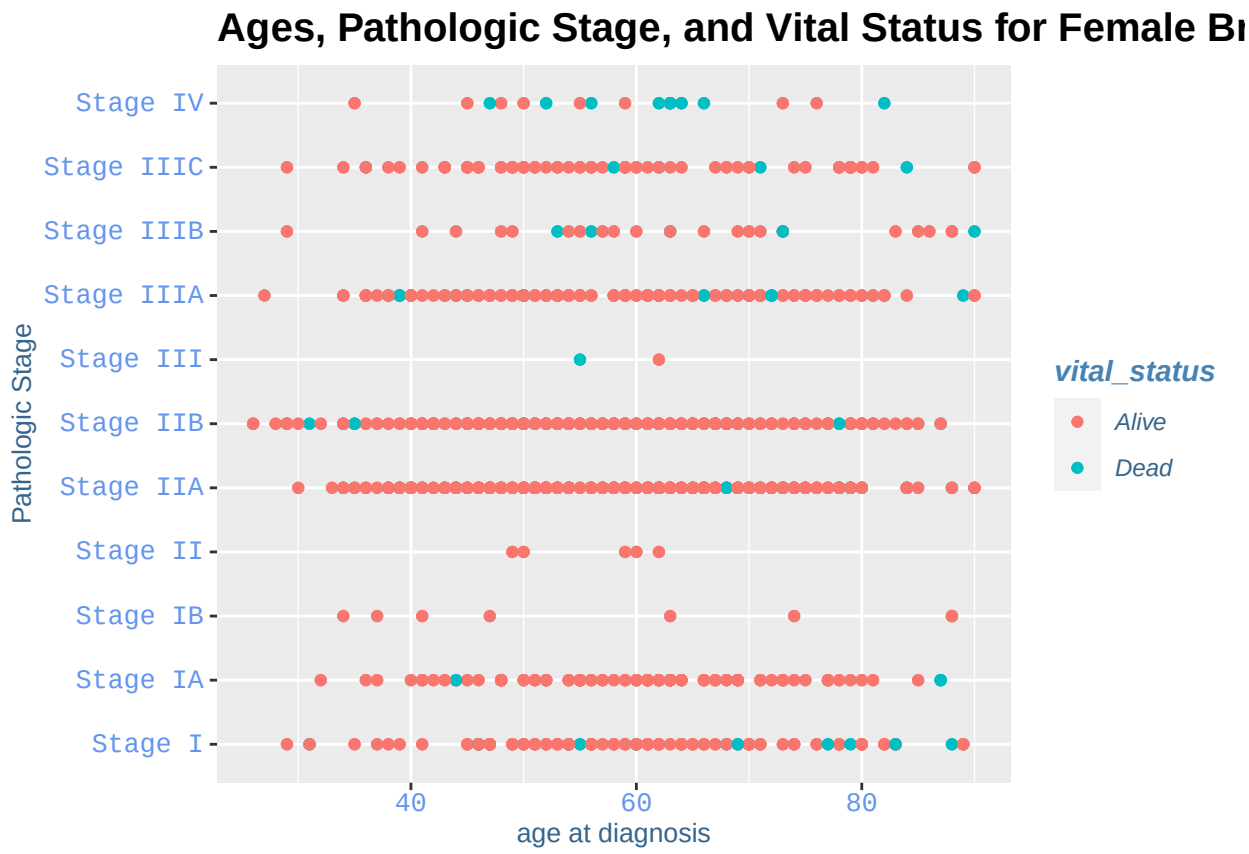

Figure 3. Scatter plot for ages, pathologic stage, and vital status for female breast cancer patients. The proportion of "dead" female patients increases with stage increases. The old female patients have more patients with "dead" vital status than young female patients.

Kaplain-Meier plot shows that the survival probability of old female patients dramatically decreases when time at 4000 days, but the young female patients' survival probability gradually increases when the time goes by at Stage I (Figure 4).

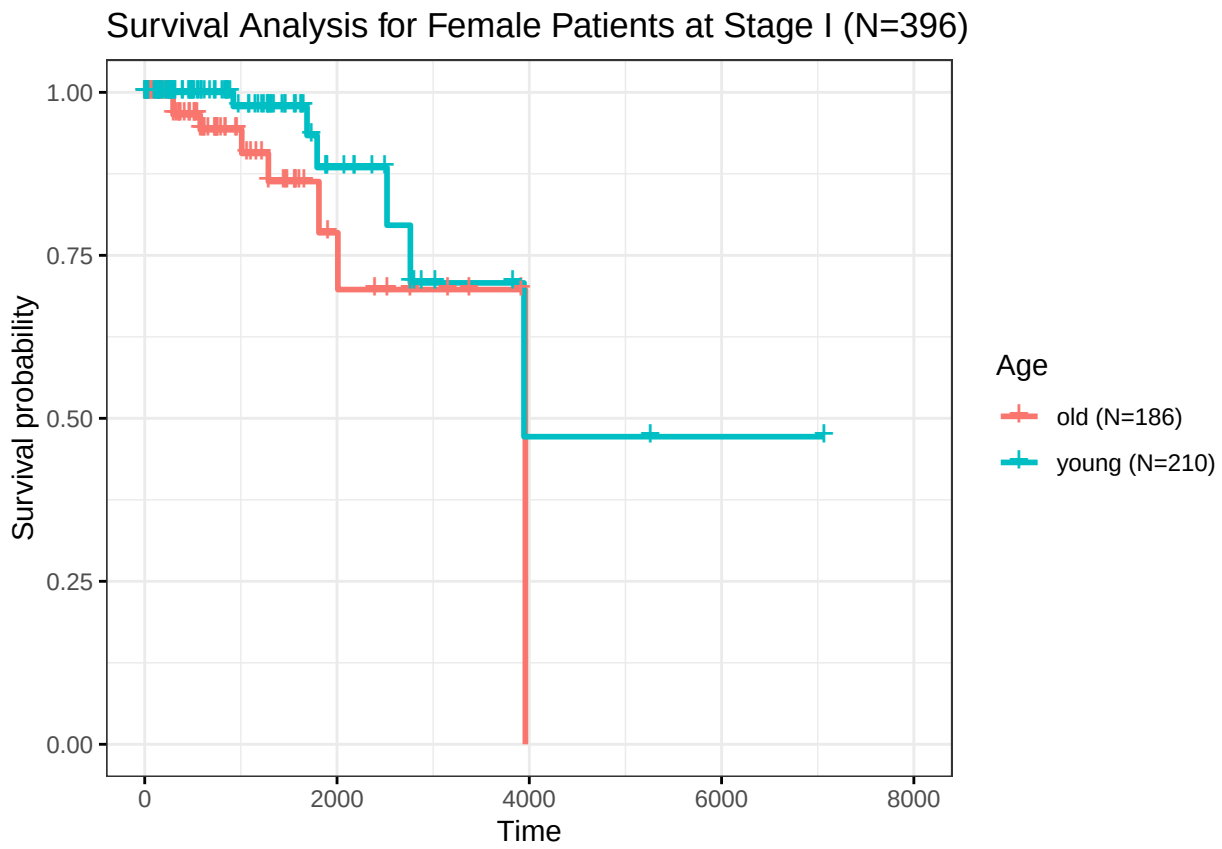

Figure 4. Kaplan-Meier plot of the survival probability between the 306 old female patients and 90 young female patients over time at Stage I. The data for this graph was taken from the patient's clinical data in the TCGA dataset. The survival probability was calculated by the vital status of the patients based on the days until death as the time. At time around 4000 days, the survival probability of old female patients dramatically decreases.

When it comes to stage II and stage III, the survival probability for young female patients is higher than old female patients with time passes. (Figure 5 and 6)

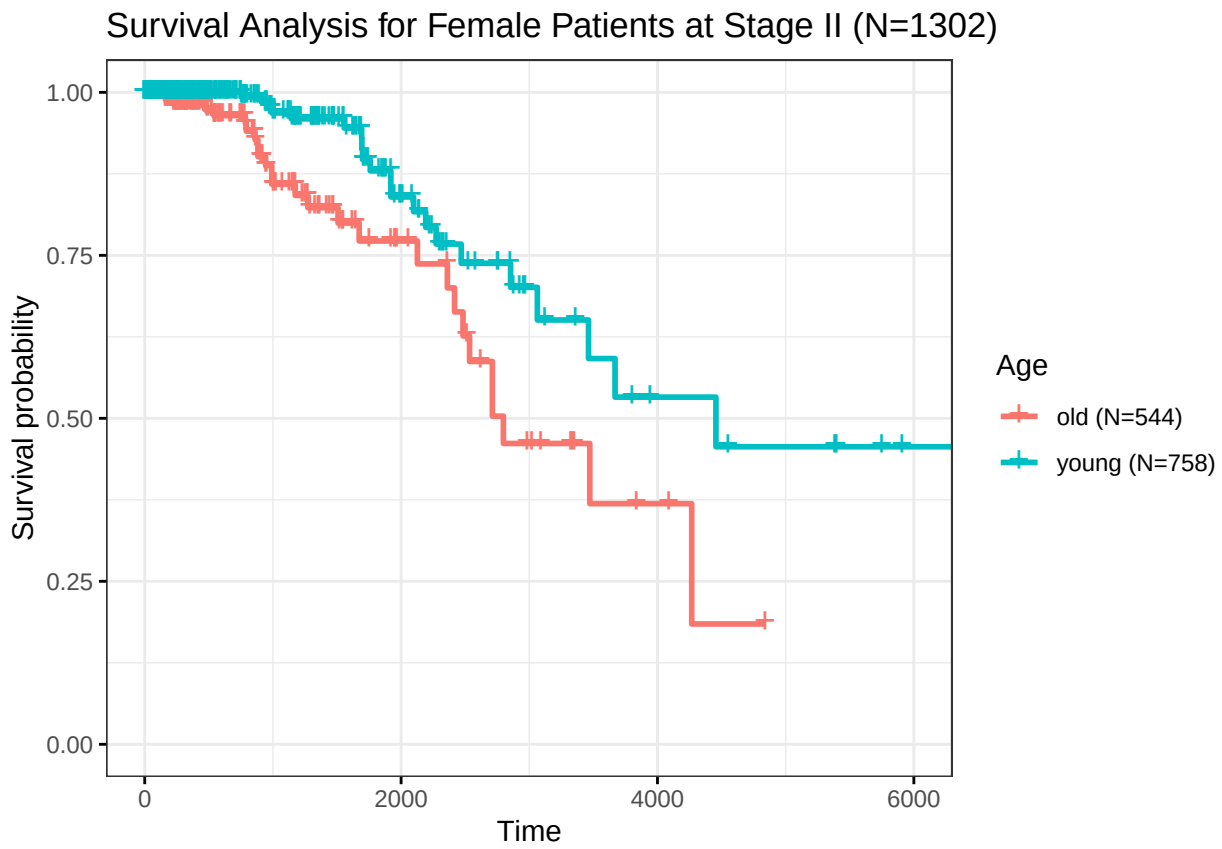

Figure 5. Kaplan-Meier plot of the survival probability between the 952 old female patients and 350 young female patients over time at Stage II. The data for this graph was taken from the patient's clinical data in the TCGA dataset. The survival probability was calculated by the vital status of the patients based on the days until death as the time. The survival probability for young female patients is higher than old female patients with time passes.

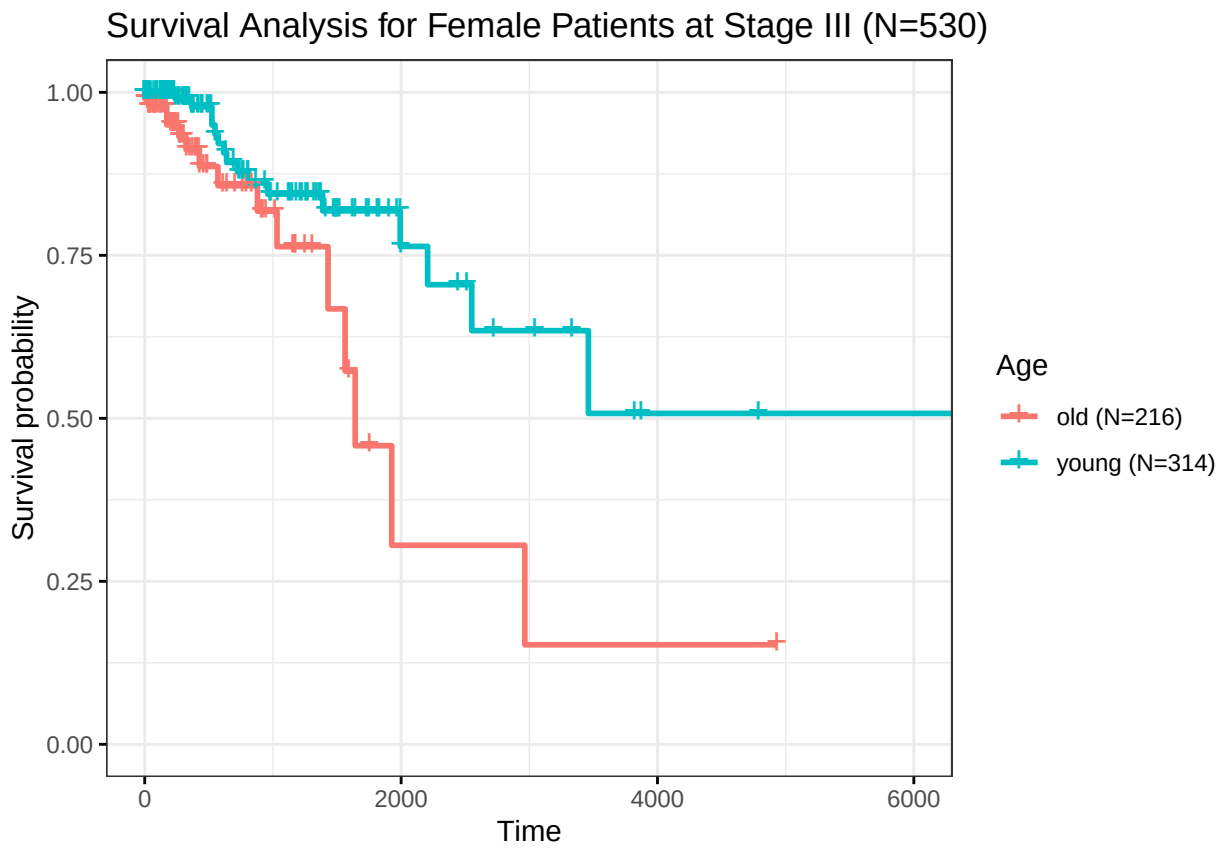

Figure 6. Kaplan-Meier plot of the survival probability between the 216 old female patients and 314 young female patients over time at Stage III. The data for this graph was taken from the patient's clinical data in the TCGA dataset. The survival probability was calculated by the vital status of the patients based on the days until death as the time. The survival probability for young female patients is higher than old female patients with time passes.

Keplain Meier Plot in Figure 7 shows that at stage IV, the survival probability of young patients dramatically decrease around time at 1000 days. In the contrast, the survival probability of old patients dramatically decreases around time goes to 4000 days.

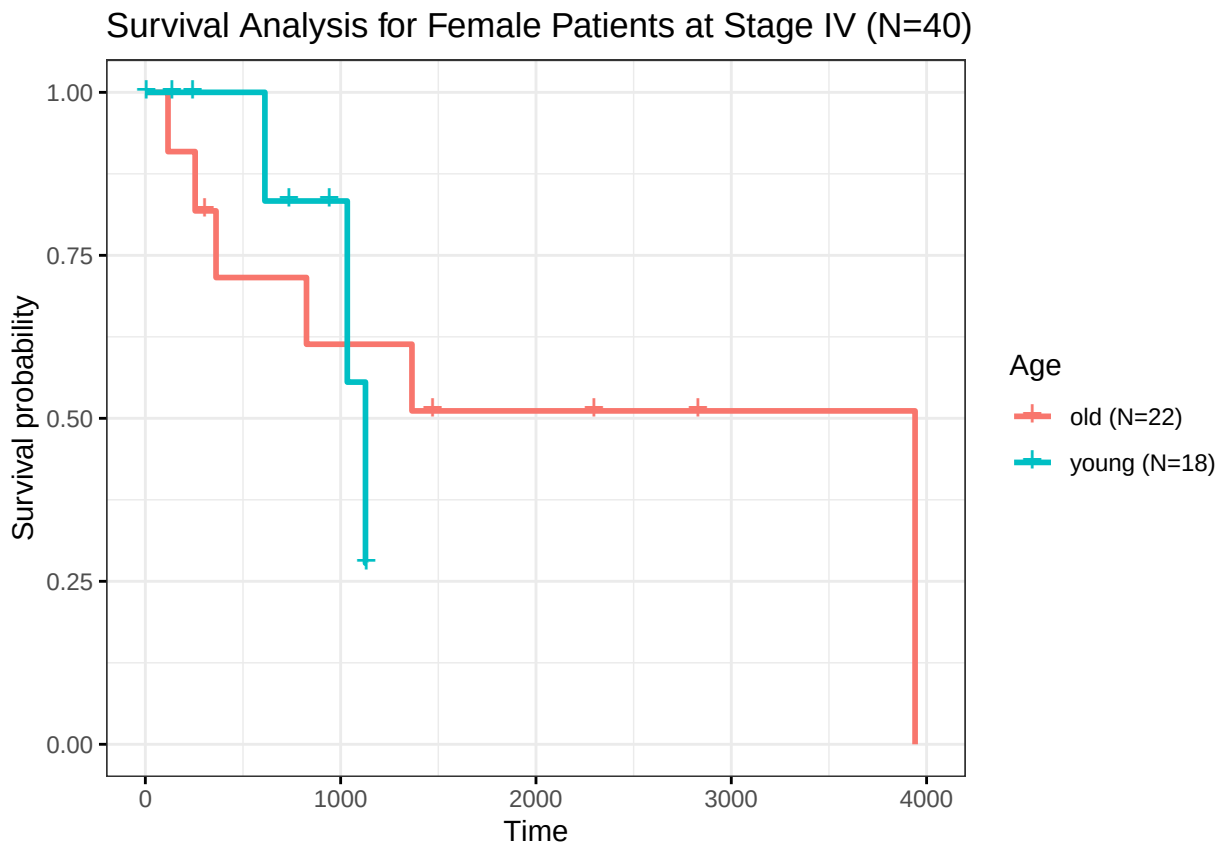

Figure 7. Kaplan-Meier plot of the survival probability between the 32 old female patients and 8 young female patients over time at Stage IV. The data for this graph was taken from the patient's clinical data in the TCGA dataset. The survival probability was calculated by the vital status of the patients based on the days until death as the time. The survival probability of young patients dramatically decrease around time at 1000 days.

Conclusion

First, we examine the structure of our clinic data with three factors: age, pathologic stage, and vital status. In general, the rate of death was higher for female patients in the higher state (for example, in stage IV). At the same time, for female patients in the same stage, the rate of death was higher for old patients. When we examine the survival probability in each stage, we cannot get a general conclusion since the population in each stage is varied. But if we see the survival analysis in stage II and stage III (with the most population), we see that the survival probability of young female patients is higher than that of old female patients.

The research still faces many limitations, including scarce sample size, and more quantitative data to diagnoze different stages, etc. For my future research, I will continue to get more data for female breast cancer patients and construct a more convincing survival analysis. Also, I want to deeply explore the reason to cause the difference in survival probability in different pathologic stage.

Reference

Breast cancer statistics: How common is breast cancer? American Cancer Society. (n.d.). Retrieved December 5, 2022, from https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html

Stage 1. Stage 1 | Breast cancer | Cancer Research UK. (2020, July 10). Retrieved December 5, 2022, from https://www.cancerresearchuk.org/about-cancer/breast-cancer/stages-types-grades/number-stages/stage-1#:~:text=Stage%201%20breast%20cancer%20means,breast%20cancer%20is%20very%20complex.

Stage 2. Stage 2 | Breast cancer | Cancer Research UK. (2020, July 10). Retrieved December 5, 2022, from https://www.cancerresearchuk.org/about-cancer/breast-cancer/stages-types-grades/number-stages/stage-2#:~:text=Stage%202%20breast%20cancer%20means,decide%20which%20treatment%20you%20need.

Stage 3. Stage 3 | Breast cancer | Cancer Research UK. (2020, July 10). Retrieved December 5, 2022, from https://www.cancerresearchuk.org/about-cancer/breast-cancer/stages-types-grades/number-stages/stage-3#:~:text=Stage%203%20means%20that%20the,decide%20which%20treatment%20you%20need.

Treatment of stage IV (metastatic) breast cancer. American Cancer Society. (n.d.). Retrieved December 5, 2022, from https://www.cancer.org/cancer/breast-cancer/treatment/treatment-of-breast-cancer-by-stage/treatment-of-stage-iv-advanced-breast-cancer.html#:~:text=Stage%20IV%20cancers%20have%20spread,the%20brain%20or