Comparing the effectiveness of risk factor about Her-2 low with fluorescence in situ hybridization breast cancer metastasis among women aged greater or equal to 45 in China.

Yi Yang

Class: GPH GU-2686 Thesis I Section 006

Dec. 18. 2023

1. **Introduction**

1. 1 Background

Breast cancer is the most common malignant tumor of women, and its mortality rate ranks second among female tumors even though the comprehensive treatment level has been constantly improving.1 Breast tumors usually start from the ductal hyperproliferation, and then develop into benign tumors or even metastatic carcinomas after constant stimulation by various carcinogenic factors.2

Her2 belongs to the Her/ErbB2/Neu family of transmembrane receptors, alongside Her1/EGFR, Her3, and Her4. These receptors are characterized by three distinct regions: an extracellular domain (ECD), a membrane-spanning region, and a cytoplasmic tyrosine kinase domain. 3 In 15-20% of breast cancers, the member of the HER receptor family, HER2, is overexpressed. 4 Although specific risk factors associated with economic development are largely unavoidable, the projected substantial rise in new cases of breast cancer underscores the urgency of incorporating this disease into future healthcare infrastructure planning at an early stage. 5 In the contemporary landscape, the increasing emphasis on Her-2 low expression breast cancer is propelled by the advancement of antibody-drug conjugates (ADC). Recent studies indicate that the Her-2 receptor is present to some extent on the cell membrane, drawing heightened attention to the exploration of its implications in breast cancer research. 6 Her-2 low is defined as Her-2 (1+) or Her-2 (2+) with ISH (-) according to the current clinical scoring system by most of the published data and ongoing clinical trials. 7–15 Several studies have reported poorer prognosis in Her-2 low positive breast cancer patients.16–19 But so far, no such results in Chinese patients are reported. In China, Her-2 low-positive breast cancers made up 54% of all patients, as calculated by a multi-center retrospective research. 5

Fluorescence in situ hybridization (FISH) serves as an alternative technique for evaluating HER2 status in clinical specimens. This method, directly quantifying the HER2 gene copy number, has proven to be a dependable approach for assessing HER2 status in formalin-fixed, paraffin-embedded clinical specimens. Its demonstrated high sensitivity and specificity rates underscore its reliability in determining the HER2 status, providing a valuable tool for accurate molecular characterization in clinical settings. 20

Presently, the identification of breast cancer metastasis hinges on several methods, including the assessment of clinical symptoms indicative of distant organ involvement, biopsies extracted from affected organs, radiological examinations, various imaging modalities, and the measurement of serum tumor markers.21,22 Furthermore, ASCO also advocates for the use of mammography as a means of early detection for relapse in breast cancer.23 The process of metastasis comprises a series of sequential steps. Failure to complete any of these steps will arrest the process. 24 Metastasis starts with the local invasion of surrounding host tissue by cells originating from the primary tumor and continues until the tumor cells invade and intravasate into blood or lymphatic vessels. 25,26 The tumor cells are disseminated via the bloodstream or the lymphatic vessels to distant organs. Consequently, the tumor cells undergo cell cycle arrest and adhere to capillary beds within the target organ, before extravasating into the organ parenchyma, proliferating and promoting angiogenesis within the organ.25 During the progression through these steps, the tumor cells must concurrently navigate the host's immune response and evade apoptotic signals to ensure their survival.25,27 Should the tumor cells successfully navigate through these steps, the process can be iterated, leading to the formation of secondary metastases, commonly referred to as 'metastasis of metastases.' 24,26

According to the recent study by Dr.Seung from Samsung Medical Center in korea has shown that Tumor size, number of Lymph nodes,and CA125 are three significant variables that are related to Her-2 low with FISH breast cancer, and these three variables are also the exposure for this study.28 Many of the studies have mentioned that bivariable analysis, multivariate analysis, and causal analysis are three important methods for data analysis in public health. The result for these data analysis can significantly Improve understanding of the disease with exposure.29,30

CA125 is an antigen associated with coelomic epithelium and serves as a valuable marker for monitoring residual disease in individuals undergoing chemotherapy for ovarian cancer. The interpretation of serum CA125 levels traditionally relies on a standard value of 35 U/ml. This benchmark was established by screening a diverse cohort of young blood donors from the general population, inclusive of women with intact reproductive systems. 31

Traditionally, the size of a breast cancer at the time of diagnosis has been regarded as a crucial factor influencing clinical outcomes. Nevertheless, the aggressive behavior exhibited by certain subtypes of breast cancer, even when they are small (≤1 cm in diameter), challenges the assumption that cancer size should be the sole consideration in treatment decisions. While an association between tumor size and lymph-node involvement holds for most tumor types, this correlation is not universally consistent. In specific subtypes of breast cancers, the relationship among tumor size, lymph-node status, and prognosis may indicate an underlying disproportionate connection between the number of cancer cells and their metastatic potential. This nuanced understanding urges a more comprehensive evaluation beyond tumor size alone when determining optimal treatment strategies for certain breast cancer subtypes. 32

1.2 Gaps and Significance

Limited understanding of how Her-2 low breast cancer with FISH (fluorescence in situ hybridization) relates to breast cancer and metastasis in elderly Chinese women is the biggest gap. While extensive research has elucidated various aspects of breast cancer, the specific characteristics and implications of Her-2 low status in the context of FISH remain insufficiently explored, especially within the demographic of elderly Chinese women. This demographic subset often faces distinct challenges related to healthcare access, comorbidities, and treatment responses. By addressing this gap, the study aims to bridge the knowledge deficit and unravel critical insights into how Her-2 low breast cancer, as identified through FISH, contributes to the metastatic process in elderly Chinese women. Such findings would not only advance the understanding of breast cancer heterogeneity but also provide essential information for tailoring targeted and personalized interventions, thereby improving the prognosis and treatment outcomes for this specific population.

The most significant result of this study will highlight the crucial roles of risk factor for Her-2 low with FISH breast cancer metastasis in this particular population. Given the rising incidence of breast cancer in China and the increasing recognition of Her-2 low as a distinct entity, the outcomes of this study may have profound implications for patient management, contributing to improved clinical decision-making and ultimately enhancing the overall quality of breast cancer care for women in this demographic group.

1. **Method**

2.1 Data Source and Analysis Tools.

The data for this study were sourced from icpsr.com, involving a retrospective cohort study conducted from May 14th, 2010, to October 26th, 2019, with a focus on 1874 breast cancer patients who underwent surgery at Beijing Pinggu Hospital in China. All patients received essential adjuvant or neoadjuvant treatments, such as chemotherapy, endocrine therapy, and radiotherapy, tailored to their pathology following the NCCN (National Comprehensive Cancer Network) breast cancer guidelines. Post-treatment, patients were scheduled for outpatient reviews every 3 months during the initial 2 years and every 6 months in the subsequent 3 years. After the completion of 5 years, surveillance reviews were conducted annually. All variables name, and corresponding values in the study are listed at the bottom of the proposal.33

In this study, R, Stata, and SAS programming languages will be used for data upload, data cleaning, and data analysis.

2.2 Research Question and Hypothesis.

The Research question for the study is How do different risk factors contribute to the metastasis of Her-2 low with FISH breast cancer, among women aged 45 or above in the Chinese population? So the outcome variable would be Metastasis, when the result is one, it means that the breast cancer has metastasized and vice versa. Exposure variables are Tumor Size, Number of Lymph nodes, CA125. In the exposure variables Tumor Size, Number of Lymph nodes, and CA125 are continuous variables.

The Hypothesis for the study is that there is no significant risk factor that can influence the metastasis rates of women who have Her-2 low with FISH breast cancer in the specified age group (≥45 years age) in China.

2.3 Statistical Analysis Plan.

The full statistical analysis process could be conducted with four important sections which are data cleaning, descriptive statistics, correlation analysis, and casual relationship analysis.

Data cleaning is a crucial and indispensable step in the data analysis and decision-making process. Raw data, often collected from various sources, is prone to errors, inconsistencies, and inaccuracies. Without proper cleaning, these flaws can propagate throughout the analysis, leading to flawed insights and unreliable conclusions. Data cleaning involves identifying and rectifying missing values, outliers, duplicates, and formatting issues, ensuring that the dataset is accurate, complete, and consistent. Clean data lays the foundation for robust statistical analyses, machine learning models, and meaningful visualizations, enabling organizations to make informed and confident decisions. Moreover, clean data enhances the efficiency of data processing, reduces the risk of errors, and fosters a trustworthy data environment. Ultimately, the integrity and reliability of any data-driven initiative hinge on the meticulous and thorough process of data cleaning.

Descriptive analysis is a fundamental approach in statistics and data analytics that involves summarizing, organizing, and presenting key characteristics of a dataset. The primary goal of descriptive analysis is to distill complex information into a more understandable and manageable form, providing insights into the essential features of the data. Descriptive analysis includes the use of graphical representations like histograms, box plots, and scatter plots to visually depict the distribution and patterns within the data. Descriptive analysis serves as an initial step in data exploration, helping analysts and researchers gain a clear understanding of the dataset's structure and trends. For example, Examining the distribution of data before analysis is imperative as it provides valuable insights into the underlying patterns and characteristics of the dataset. Understanding the distribution helps researchers and analysts make informed decisions about the most appropriate statistical methods to apply. Different analysis techniques assume specific data distributions, and deviations from these assumptions can lead to biased or inaccurate results. By scrutinizing the distribution, analysts can identify outliers, assess skewness, and choose appropriate measures of central tendency and dispersion. This preliminary step not only ensures the validity of statistical tests but also aids in selecting suitable visualization techniques that effectively represent the data. Additionally, knowledge of the data distribution allows for the implementation of appropriate data transformation techniques if needed. In essence, checking the distribution before analysis is a fundamental aspect of robust data exploration, facilitating the accurate interpretation of results and enhancing the overall reliability of research findings.

Multivariate analysis plays a pivotal role in data-driven decision-making by allowing researchers and analysts to explore the intricate relationships among multiple variables simultaneously. Unlike univariate or bivariate analyses, which focus on individual or paired variables, multivariate analysis provides a comprehensive understanding of complex systems where variables interact and influence each other. This holistic approach is particularly valuable in fields such as statistics, data science, and social sciences, where the interdependencies among various factors are crucial for a nuanced understanding of phenomena. Multivariate analysis facilitates dimensionality reduction, enabling researchers to distill essential information from large datasets, and it supports pattern recognition to identify meaningful trends and clusters within the data. Moreover, multivariate techniques contribute to predictive modeling, optimization, and decision-making processes by considering the joint effects of multiple factors. The application of multivariate analysis enhances the robustness of findings, aids in visualizing complex datasets, and is instrumental in addressing challenges across diverse domains, from epidemiology and health research. For instance, Binomial logistic regression is specifically designed for situations where the outcome variable is binariy. By applying all exposures and outcome variables into the logistic regression model, if one of the variable’s p-value is less than 0.05, we can reject the null hypothesis, which there is at least one risk factor that is significant to influence the metastasis rates of women who have Her-2 low with FISH breast cancer in the specified age group (≥45) in China.

Causal relationship analysis is crucial in the field of data science and research because it goes beyond correlation to explore the cause-and-effect connections between variables. While correlation identifies associations, causation delves into understanding whether changes in one variable directly influence changes in another. Establishing causal relationships is fundamental for making informed decisions and developing effective strategies. By employing experimental design, statistical modeling, and advanced analytical techniques, researchers can discern the true impact of specific factors on outcomes. This type of analysis is especially valuable in fields such as public health, economics, and social sciences, where understanding the root causes of phenomena is paramount. Causal relationship analysis not only provides deeper insights into the mechanisms driving observed patterns but also aids in predicting and controlling outcomes. It is essential for constructing robust models and interventions, enabling practitioners and decision-makers to implement evidence-based solutions that address the underlying causes rather than merely addressing surface-level correlations. In essence, causal relationship analysis is a key step towards achieving a more comprehensive and nuanced understanding of complex systems.

1. **Limitations and Potential Bias**

The primary bias in this study is rooted in the absence of comprehensive demographic data, which poses a significant limitation to the overall validity and reliability of the findings. Without a thorough understanding of the diverse characteristics of the study population, the results may not accurately represent the broader demographic landscape. This lack of inclusivity not only hampers the generalizability of the study but also raises ethical concerns, as it may inadvertently perpetuate disparities or overlook important nuances within specific subgroups. Moreover, the narrow demographic scope further compounds these issues, limiting the study's applicability to a more diverse or representative population. Consequently, addressing these biases becomes imperative to ensure the study's ethical soundness and enhance the external validity of its conclusions.

A significant methodological limitation in this study revolves around the dependence on the quality and quantity of the data used. The reliability and validity of the study's results are intricately tied to the accuracy and completeness of the data collected, introducing a potential source of bias and uncertainty. Moreover, the study's ability to generalize findings to new, unseen data remains uncertain, as the performance on unfamiliar datasets is not guaranteed. This limitation underscores the importance of cautious interpretation and generalization of the study's outcomes, emphasizing the need for robust validation measures and consideration of potential biases stemming from data-related challenges.

1. **Timelines**

|  |  |  |
| --- | --- | --- |
| Week | Start Date | Tasks |
| Week 1 | Jan, 15, 2024 | Meet with thesis advisor on revised thesis proposal. |
| Week 2 | Jan 22, 2024 | Write introduction and discussion. |
| Week 3 | Jan, 29, 2024 | Attain de-identified data, clean dataset. |
| Week 4 | Feb, 5, 2024 | Clean dataset, get suggestions on statistical analysis, and Sensitivity analysis to test the robustness (Outliers). |
| Week 5 | Feb, 12, 2024 | Produce Logistic Regression, and interpret the correlation result. |
| Week 6 | Frb, 19, 2024 | Get Feedback on Logistic Regression results, start Univariable analysis for selected variables, and interpret the results. |
| Week 7 | Feb, 26, 2024 | Get Feedback on Univariable analysis results, start Bivariable analysis for selected variables, and interpret the results. |
| Week 8 | Mar, 4, 2024 | Get Feedback on Bivariable analysis result, and start check for multicollinearity among independent variables |
| Week 9 | Mar, 11,2024 | Get Feedback on multicollinearity analysis results, and Stratification for difference in subgroups if needed. |
| Week 10 | Mar, 18, 2024 | Pull together all sections and revise; add citations |
| Week 11 - 12 | Mar, 25, 2024 | Get help on proofreading and editing |
| Week 13 | Apr, 3, 2024 | Submit Final Project |

Reference

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71:209–49.
2. Sun Y-S, Zhao Z, Yang Z-N, et al. Risk factors and preventions of breast cancer. *International Journal of Biological Sciences*. 2017;13(11):1387-1397. doi:10.7150/ijbs.21635
3. Windy Dean-Colomb, Francisco J. Esteva. Her2-positive breast cancer: Herceptin and beyond, European Journal of Cancer. Volume 44. Issue 18. 2008. Pages 2806-2812. ISSN 0959-8049,<https://doi.org/10.1016/j.ejca.2008.09.013>.
4. Sapino A, Goia M, Recupero D, Marchiò C. Current Challenges for HER2 Testing in Diagnostic Pathology: State of the Art and Controversial Issues. Front Oncol. 2013;3:129. Published 2013 May 21. doi:10.3389/fonc.2013.00129.
5. Shui R, Liang X, Li X, et al. Hormone Receptor and Human Epidermal Growth Factor Receptor 2 Detection in Invasive Breast Carcinoma: A Retrospective Study of 12,467 Patients From 19 Chinese Representative Clinical Centers. Clin Breast Cancer. 2020;20(1):e65-e74. doi:10.1016/j.clbc.2019.07.013
6. Onsum MD, Geretti E, Paragas V, et al. Single-cell quantitative HER2 measurement identifies heterogeneity and distinct subgroups within traditionally defined HER2-positive patients. *The American Journal of Pathology*. 2013;183(5):1446-1460. doi:10.1016/j.ajpath.2013.07.015
7. Banerji U, van Herpen CM, Saura C, et al. Trastuzumab duocarmazine in locally advanced and metastatic solid tumours and HER2-expressing breast cancer: A phase 1 dose-escalation and dose-expansion study. *The Lancet Oncology*. 2019;20(8):1124-1135. doi:10.1016/s1470-2045(19)30328-6
8. Chick RC, Clifton GT, Hale DF, et al. Subgroup analysis of Nelipepimut-S Plus GM-CSF combined with trastuzumab versus trastuzumab alone to prevent recurrences in patients with high-risk, HER2 low-expressing breast cancer. *Clinical Immunology*. 2021;225:108679. doi:10.1016/j.clim.2021.108679
9. Mittendorf EA, Lu B, Melisko M, et al. Efficacy and safety analysis of Nelipepimut-s vaccine to prevent breast cancer recurrence: A randomized, multicenter, phase III clinical trial. *Clinical Cancer Research*. 2019;25(14):4248-4254. doi:10.1158/1078-0432.ccr-18-2867
10. Modi S, Saura C, Yamashita T, et al. Trastuzumab Deruxtecan in previously treated her2-positive breast cancer. *New England Journal of Medicine*. 2020;382(7):610-621. doi:10.1056/nejmoa1914510
11. Schneeweiss A, Park-Simon T-W, Albanell J, et al. Phase IB study evaluating safety and clinical activity of the anti-HER3 antibody lumretuzumab combined with the anti-HER2 antibody pertuzumab and paclitaxel in her3-positive, HER2-low metastatic breast cancer. *Investigational New Drugs*. 2018;36(5):848-859. doi:10.1007/s10637-018-0562-4
12. Clifton GT, Peace KM, Holmes JP, et al. Initial safety analysis of a randomized phase II trial of nelipepimut-S + GM-CSF and trastuzumab compared to trastuzumab alone to prevent recurrence in breast cancer patients with HER2 low-expressing tumors. *Clinical Immunology*. 2019;201:48-54. doi:10.1016/j.clim.2019.02.011
13. Pondé N, Aftimos P, Piccart M. Antibody-Drug Conjugates in Breast Cancer: a Comprehensive Review. *Curr Treat Options Oncol*. 2019;20(5):37. Published 2019 Apr 1. doi:10.1007/s11864-019-0633-6
14. Rinnerthaler G, Gampenrieder SP, Greil R. HER2 Directed Antibody-Drug-Conjugates beyond T-DM1 in Breast Cancer. *Int J Mol Sci*. 2019;20(5):1115. Published 2019 Mar 5. doi:10.3390/ijms20051115
15. Modi S, Jacot W, Yamashita T, et al. Trastuzumab Deruxtecan in previously treated HER2-low advanced breast cancer. *New England Journal of Medicine*. 2022;387(1):9-20. doi:10.1056/nejmoa2203690
16. Camp RL, Dolled-Filhart M, King BL, Rimm DL. Quantitative analysis of breast cancer tissue microarrays shows that both high and normal levels of HER2 expression are associated with poor outcome. *Cancer Res*. 2003;63(7):1445-1448.
17. Gilcrease MZ, Woodward WA, Nicolas MM, et al. Even low-level HER2 expression may be associated with worse outcome in node-positive breast cancer. *Am J Surg Pathol*. 2009;33(5):759-767. doi:10.1097/PAS.0b013e31819437f9
18. Eggemann H, Ignatov T, Burger E, et al. Moderate HER2 expression as a prognostic factor in hormone receptor positive breast cancer. *Endocr Relat Cancer*. 2015;22(5):725-733. doi:10.1530/ERC-15-0335
19. Rossi V, Sarotto I, Maggiorotto F, et al. Moderate immunohistochemical expression of HER-2 (2+) without HER-2 gene amplification is a negative prognostic factor in early breast cancer. *Oncologist*. 2012;17(11):1418-1425. doi:10.1634/theoncologist.2012-0194
20. Pauletti G, Godolphin W, Press MF, Slamon DJ. Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization. *Oncogene*. 1996;13(1):63-72.
21. Lacroix M. Significance, detection and markers of disseminated breast cancer cells. *Endocrine-Related Cancer*. 2006;13(4):1033-1067. doi:10.1677/erc-06-0001
22. Sun Y-F, Yang X-R, Zhou J, Qiu S-J, Fan J, Xu Y. Circulating tumor cells: Advances in detection methods, biological issues, and clinical relevance. *Journal of Cancer Research and Clinical Oncology*. 2011;137(8):1151-1173. doi:10.1007/s00432-011-0988-y
23. Khatcheressian JL, Wolff AC, Smith TJ, et al. American Society of Clinical Oncology 2006 update of the breast cancer follow-up and management guidelines in the adjuvant setting. *Journal of Clinical Oncology*. 2006;24(31):5091-5097. doi:10.1200/jco.2006.08.8575
24. Poste G, Fidler IJ. The pathogenesis of cancer metastasis. *Nature*. 1980;283(5743):139-146. doi:10.1038/283139a0
25. Hunter KW, Crawford NP, Alsarraj J. Mechanisms of metastasis. *Breast Cancer Research*. 2008;10(S1). doi:10.1186/bcr1988
26. Talmadge JE, Fidler IJ. AACR Centennial Series: The Biology of Cancer Metastasis: Historical Perspective. *Cancer Research*. 2010;70(14):5649-5669. doi:10.1158/0008-5472.can-10-1040
27. Fidler IJ, Gersten DM, Hart IR. The biology of cancer invasion and metastasis. *Adv Cancer Res*. 1978;28:149-250. doi:10.1016/s0065-230x(08)60648-x
28. Min SK, Lee SK, Woo J, et al. Relation Between Tumor Size and Lymph Node Metastasis According to Subtypes of Breast Cancer. *J Breast Cancer*. 2021;24(1):75-84. doi:10.4048/jbc.2021.24.e4
29. Fountain FF, Tolley E, Chrisman CR, Self TH. Isoniazid hepatotoxicity associated with treatment of latent tuberculosis infection. *Chest*. 2005;128(1):116-123. doi:10.1378/chest.128.1.116
30. Ong SK, Haruyama R, Yip CH, et al. Feasibility of Monitoring Global Breast Cancer Initiative Framework Key Performance Indicators in 21 Asian National Cancer Centers Alliance member countries. *eClinicalMedicine*. 2024;67:102365. doi:10.1016/j.eclinm.2023.102365
31. Alagoz T, Buller RE, Berman M, Anderson B, Manetta A, DiSaia P. What is a normal CA125 level? *Gynecologic Oncology*. 1994;53(1):93-97. doi:10.1006/gyno.1994.1093
32. Foulkes WD, Reis-Filho JS, Narod SA. Tumor size and survival in breast cancer—a reappraisal. *Nature Reviews Clinical Oncology*. 2010;7(6):348-353. doi:10.1038/nrclinonc.2010.39
33. Li M, Li X, Xu F, Jiang H. *A retrospective 10 year survival analysis of 1340 HER-2 low positive breast cancer patients*. Published online 2023. doi:10.21203/rs.3.rs-3028676/v1

**Variable List**:

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Responses** | **Description** |
| Age | Numeric Variable | Patient Age. |
| CA125 | Numeric Variable | The CA125 blood test measures the level of this protein in the blood. |
| CA15-3 | Numeric Variable | CA15-3 is a protein that can be measured through a blood test to help monitor and manage breast cancer. |
| Interval (surgery - most recent examination) | Numeric Variable | The Interval time for the most recent examination. |
| Surgical | 1=Breast-conserving modified radical treatment;  2=Puncture biopsy;  3=simple resection;  4=Improved and radical cure;  5=radical surgery;  6=local expansion and cutting;  7=tumor resection | What Type of Surgical that patient received.  1 = Breast-conserving modified radical treatment: This likely refers to a surgical approach that removes the tumor and some surrounding healthy tissue while preserving as much of the breast as possible. This is often done in combination with radiation therapy.  2 = Puncture biopsy: This is a diagnostic procedure in which a small sample of tissue is removed from the breast using a needle for examination under a microscope. It is commonly used for obtaining a tissue sample for diagnosis.  3 = Simple resection: This suggests a surgical procedure where a portion of the breast tissue containing the tumor is removed. It may be less extensive than a modified radical treatment.  4 = Improved and radical cure: The term "improved and radical cure" is not standard surgical terminology. "Radical cure" might suggest a surgery with the intent of completely removing the cancer.  5 = Radical surgery: This refers to a more extensive surgical procedure aimed at removing the entire tumor and possibly surrounding tissues or lymph nodes. It is typically performed when the cancer has spread or is more aggressive.  6 = Local expansion and cutting: This term is not commonly used in standard surgical descriptions. It might refer to procedures involving the expansion or removal of tissues in the localized area of the tumor.  7 = Tumor resection: This is a general term indicating the surgical removal of a tumor. The extent of resection can vary from a partial removal to complete removal, depending on the surgical approach. |
| Histological type | 1=Metaplastic carcinoma,  2=Invasive ductal carcinoma,  3=Invasive basal-like carcinoma,  4=Invasive cribriform carcinoma,  5=Invasive lobular carcinoma,  6=Invasive cancer,  7=Papillary carcinoma,  8=Neuroendocrine carcinoma,  9=Medullary carcinoma,  10=micropapillary carcinoma,  11=Undetermined. | What kind of Histological type that patient has.  1 = Metaplastic carcinoma: A rare type of breast cancer where the cancer cells have characteristics of both epithelial (glandular) and mesenchymal (connective tissue) cells.  2 = Invasive ductal carcinoma: The most common type of breast cancer, where cancer cells invade the surrounding breast tissues from the milk ducts.  3 = Invasive basal-like carcinoma: A subtype of invasive ductal carcinoma with characteristics similar to the basal cells of the breast.  4 = Invasive cribriform carcinoma: A type of breast cancer where the cancer cells form a Swiss cheese-like pattern when viewed under a microscope.  5 = Invasive lobular carcinoma: A type of breast cancer that begins in the milk-producing glands (lobules) and invades nearby tissues.  6 = Invasive cancer: A general term indicating cancer that has invaded surrounding tissues.  7 = Papillary carcinoma: A type of breast cancer where tumor cells form small, finger-like projections.  8 = Neuroendocrine carcinoma: A type of breast cancer with features similar to neuroendocrine cells, which produce hormones.  9 = Medullary carcinoma: A type of breast cancer with well-defined borders and a lymphocyte-rich appearance.  10 = Micropapillary carcinoma: A subtype of invasive ductal carcinoma characterized by small, finger-like projections.  11 = Undetermined: This category is used when the histological type cannot be definitively determined or if additional tests are needed for accurate classification. |
| Neural invasion | 0 = No,  1 = Yes. | Neural invasion refers to the infiltration or penetration of nerves by cancer cells. |
| Lymphatic or blood vascular tumor emboli | 0 = No,  1 = Yes. | Lymphatic or blood vascular tumor emboli refer to the spread of cancer cells through the lymphatic or blood vessels. |
| > 2cm or not | 0 = No,  1 = Yes. | Check if the patient's tumor size is greater than 2 cm or not. |
| T Stage | 1=T  2=T2  3=T3  4=T4 | T stage refers to the extent of the primary tumor in the staging of cancer, particularly in the context of the TNM staging system.  1 = T (Tumor): This stage typically indicates the primary tumor. It suggests that there is a tumor present but does not provide specific information about its size or extent.  2 = T2: This stage suggests that the primary tumor is larger or more extensive than in the T stage. The specific criteria for T2 can vary depending on the type of cancer.  3 = T3: This stage indicates a further increase in the size or extent of the primary tumor compared to T2. Again, the specific criteria for T3 depend on the cancer type.  4 = T4: This stage suggests a larger or more invasive primary tumor than T3. T4 tumors are often associated with more advanced disease. |
| N Stage | 1=N1  2=N2  3=N3 | The N stage in cancer staging refers to the involvement of regional lymph nodes.  1 = N1: This stage indicates the involvement of regional lymph nodes. The specific criteria for N1 can vary depending on the type of cancer, but generally, it means that cancer has spread to nearby lymph nodes.  2 = N2: This stage suggests further involvement of regional lymph nodes compared to N1. The criteria for N2 depend on the specific characteristics of the cancer, but it typically signifies more extensive lymph node involvement.  3 = N3: This stage indicates even more extensive involvement of regional lymph nodes compared to N2. The criteria for N3 depend on the cancer type but generally indicate a higher degree of lymph node spread. |
| TNM Stage | 1=I  2=II  3=III  4=IV | The TNM staging system is a widely used method for describing the extent of cancer based on three key factors: the size and extent of the primary tumor (T), the involvement of regional lymph nodes (N), and the presence of distant metastasis (M).  1 = Stage I (TNM I): This stage generally indicates that the cancer is small and localized. It means the primary tumor is relatively small, and there is minimal or no lymph node involvement or metastasis to other organs.  2 = Stage II (TNM II): This stage suggests a larger primary tumor or more extensive lymph node involvement than Stage I. However, the cancer is still typically confined to the original site and has not spread to distant organs.  3 = Stage III (TNM III): This stage indicates more extensive local involvement of the primary tumor and may involve significant lymph node spread. The cancer may have grown into nearby structures or tissues.  4 = Stage IV (TNM IV): This stage signifies that cancer has spread beyond the original site to distant organs or distant lymph nodes. Stage IV is often referred to as metastatic cancer, indicating that it has spread to other parts of the body. |
| Tumor Size | Numeric Variable | The size of Tumor(cm). |
| Number of Lymph nodes | Numeric Variable | The number of lymph nodes refers to the quantity of lymph nodes that are involved or examined in the context of cancer staging. |
| ER | 1=Positive,  2=negative,  3=Unknown. | Estrogen receptors(ER) are proteins found on the surface of breast cells.  1 = Positive: Estrogen receptor-positive breast cancers are often responsive to hormone therapy, which works by blocking the effects of estrogen or reducing estrogen levels in the body.  2 = Negative: Estrogen receptor-negative breast cancers do not respond to hormone therapy in the same way as ER-positive cancers, and other treatment approaches such as chemotherapy may be considered.  3 = Unknown: This category is used when there is uncertainty about the estrogen receptor status. |
| PR | 1=Positive,  2=negative,  3=Unknown. | Progesterone receptors are proteins found on the surface of breast cells:  Positive: PR-positive breast cancers respond to hormone therapy. Hormone therapy for PR-positive breast cancer may include drugs that block the effects of progesterone or reduce progesterone levels in the body.  Negative: PR-negative cancers do not typically respond to hormone therapy targeting progesterone receptors, and other treatment approaches, such as chemotherapy, may be considered. |
| HER2 | 1=1+,  2=2+,  3=3+,  4=Positive,  5=Unknown. | HER2 status in breast cancer based on immunohistochemistry (IHC).  1+ (IHC 1+): Weak or no staining observed in the tumor cells. This is considered HER2-negative.  2+ (IHC 2+): Moderate staining observed in the tumor cells. This result is equivocal, and further testing, such as fluorescence in situ hybridization (FISH), may be recommended to determine HER2 status definitively.  3+ (IHC 3+): Strong staining observed in the tumor cells. This is considered HER2-positive.  Positive: This likely encompasses both IHC 2+ and IHC 3+ cases, indicating HER2 positivity.  Unknown: This category is used when there is uncertainty about the HER2 status. |
| Molecular typing | 1=Luminal A,  2=Luminal B,  3=HER2,  4=All Negative,  5=Unknown. | Breast cancer subtypes based on the expression of certain molecular markers:  1= Luminal A tumors often have a lower proliferation rate and a generally better prognosis.  2 = Luminal B tumors can be further categorized into Luminal B HER2-negative (HER2-) or Luminal B HER2-positive (HER2+), depending on the HER2 status.  3 = HER2-positive breast cancers tend to be more aggressive, but targeted therapies such as trastuzumab (Herceptin) have significantly improved outcomes for patients with this subtype.  4 = Triple-negative breast cancer may be more challenging to treat with hormone or HER2-targeted therapies.  5 = Unknown: This category is used when there is uncertainty about the molecular subtype of breast cancer. |
| Ki67 | Numeric Variable | The Ki-67 index is a measure of the proportion of cancer cells that are actively dividing or in the process of cell division. |
| Metastasis | 1=True  0=False | Check whether a patient has breast cancer already metastasis or not. If the cancer has metastasized then Metastasis equal to 1, otherwise leave as an empty value. |
| recur | 1=True  0=False | Check whether a patient has breast cancer recur or not. |