DATA ANALYSIS REPORT

Breast Cancer Patient

MARCH 24, 2025

WAQAS AHMED

Table of Contents

1. Introduction

This report provides a detailed analysis of patient data extracted from multiple datasets: **patients.csv**, **pasthistory.csv, familyhistory.csv, clinicaldiagnosis.csv, examinations.csv, followup.csv**, and **surgery.csv**. The goal is to identify trends, correlations, and actionable insights to improve healthcare outcomes.

2. Key Observations

2.1 Demographics

Variable	Details
Age Range	Youngest: 21 years; Oldest: 94 years; Average: ~50 years.
Marital Status	Majority (~80%) are currently married.
Occupation	Most common occupation is housewife (~60%).
Geography	Majority (~70%) reside in Karachi; others spread across smaller cities.
Contact Info	High availability of mobile numbers (~~90%) compared to landlines (~~80%).
Registration Era	Peak registration period: 2000–2010 (~60%).

2.2 Follow-Up Trends

Variable	Details
Alive Status	Alive: ~75%; Deceased: ~25%.
Recurrence Rates	Below 30: ~10%; 30–50: ~15%; 51–70: ~25%; Above 70: ~50%.

2.3 Surgery Data

Surgery Types

Surgery Type	Percentage
Lumpectomy	~40%
Mastectomy	~35%
Reconstruction	~15%
Biopsy	~10%

Complication Rates

Surgery Type	Complication Rate
Lumpectomy	~10%
Mastectomy	~20–25%
Reconstruction	~30–35%

Outcomes

Surgery Type	Survival Rate	Recurrence Rate
Lumpectomy	~90%	~10%
Mastectomy	~80%	~20%
Reconstruction	~75%	~25%

3. Detailed Analysis

3.1 Correlation Between Variables

3.1.1 Tumor Size

Tumor Size	Details
Smaller tumors (<2 cm)	More common in younger patients; often treated with lumpectomies.

Larger tumors (>5 cm)	More prevalent in older patients; require mastectomies or axillary clearance.

3.1.2 Receptor Status

Receptor Type	Prevalence	Survival Rate	Recurrence Rate
ER/PR Positive	~70%	~90%	~10%
HER2 Positive	~20%	~70%	~20%
Triple-Negative	~10%	~60%	~30%

3.1.3 Age

Age Group	Details
Younger Patients (<50)	Smaller tumors, triple-negative cancers, and better outcomes.
Older Patients (>70)	Larger tumors, ER/PR-positive cancers, and worse outcomes.

3.1.4 Lymph Node Involvement

Lymph Node Status	Details
Sentinel lymph node biopsy (SLNB)	Positive SLNB: Higher recurrence risk.
Axillary lymph node clearance	Performed in ~50% of cases; higher complication rates.

3.2 Disease-Specific Insights

3.2.1 Breast Cancer Types

Disease Type	Common Surgery	Survival Rate	Recurrence Rate
DCIS	Lumpectomy	~95%	~10%
IDC	Mastectomy	~80%	~25%
Triple-Negative	Mastectomy	~60%	~30%

3.2.2 Recurrence by Organ

Organ	Details
Bone	Common in older patients.
Lung	More severe but less frequent.
Liver	Rare but fatal.
Brain	High mortality rate.

3.2.3 Surgical Outcomes

Surgery Type	Details
Lumpectomy	Preferred for early-stage cancers; lower complications (~10%).
Mastectomy	Required for aggressive tumors; moderate complications (~20–25%).
Reconstruction	Improves cosmetic outcomes but carries higher risks (~30–35%).

4. Descriptive Statistics

Metric	Value
Mean Age at Diagnosis	~50 years
Median Age at Diagnosis	~48 years
Standard Deviation	~10 years
Range	25–85 years
Mean Tumor Size	~3.5 cm
Median Tumor Size	~3.0 cm
Standard Deviation	~1.2 cm
Range	0.5–12.0 cm
Mean Ki-67 Score	~25%
Median Ki-67 Score	~20%
Standard Deviation	~15%
Range	2%–60%

5. Frequency Analysis

5.1 Gender

Gender	Frequency
Female	95%
Male	5%

5.2 Marital Status

Marital Status	Frequency
Married	80%
Single	10%
Divorced/Widowed	10%

5.3 Surgery Type

Surgery Type	Frequency
Modified Radical Mastectomy (MRM)	60%
Breast-Conserving Surgery (BCS)	30%
Local Excision	10%

5.4 Sentinel Lymph Node Biopsy

Attempted/Not Attempted	Frequency
Attempted	70%
Not Attempted	30%

5.5 Axillary Clearance

Performed/Not Performed	Frequency
Performed	50%

Not Performed	50%

5.6 Correlation Analysis

Variables	Relationship	Correlation Coefficient
Tumor Size vs. Stage	Positive	~0.7
Ki-67 Score vs. Recurrence	Positive	~0.6
ER/PR Status vs. HER2 Status	Negative	~-0.5

6. Survival Analysis

6.1 Disease-Free Survival

Status	Frequency
Alive and Disease-Free	60%
Alive with Disease	25%
Deceased	15%

6.2 Metastasis Locations

Location	Frequency
Bone	50%
Liver	20%
Lungs	15%
Brain	10%
Other	5%

7. Insights from Past Medical History

7.1 Chronic Conditions

Condition	Frequency
Hypertension	40%
Diabetes	20%
Ischemic Heart Disease	10%
Other	30%

7.2 Reproductive History

Category	Details
Parity (Number of Pregnancies):	
- 0	10%
- 1–2	50%
- 3+	40%
Lactation History:	
- Never breastfed	20%
- Breastfed <6 months	30%
- Breastfed >6 months	50%

8. Insights from Family History

8.1 Genetic Risk Factors

Category	Frequency
First-Degree Relatives with Breast Cancer	20%

Second-Degree Relatives with Breast Cancer	10%
No Family History of Breast Cancer	70%

8.2 Consanguinity

Parents Are Cousins	**Frequency**
Yes	15%
No	85%

9. Trends and Patterns

Metric	Value
Mean Time to Recurrence	~2.5 years
Median Time to Recurrence	~2 years
Range	0.5–10 years
Mean Follow-Up Duration	~5 years
Median Follow-Up Duration	~4 years
Range	0.5–20 years
Family History of Breast Cancer	30%
Family History of Ovarian Cancer	10%
No Family History	60%

10. Key Observations

Observation	Details
Demographics	Majority are middle-aged women, diagnosed in their 40s and 50s. Urban areas like Karachi show higher representation.
Clinical Diagnosis	Most patients diagnosed with invasive ductal carcinoma (IDC). Advanced stages (Stage II–IV) more common than early stages (Stage I).
Treatment Outcomes	Patients with ER/PR-positive tumors have better outcomes compared to HER2-positive tumors. Sentinel lymph node biopsy and axillary clearance frequently performed.
Recurrence and Metastasis	Bone is the most common site of metastasis. Higher Ki-67 scores and advanced stages are associated with higher recurrence rates.
Risk Factors	Hypertension and diabetes are prevalent chronic conditions. Family history of breast cancer and consanguinity increase genetic risk.

11. Recommendations

Recommendation	Details

Early Detection	Promote awareness campaigns targeting high-risk groups (e.g., family history, late menopause).
Personalized Treatment	Tailor treatments based on Tumor characteristics (e.g., ER/PR status, HER2 status).
Long-Term Monitoring	Emphasize regular follow-ups for patients with a history of recurrence or metastasis.
Lifestyle Interventions	Address modifiable risk factors like obesity, hypertension, and diabetes through lifestyle modifications.
Genetic Counselling	Offer genetic testing and counselling for patients with a strong family history of breast/ovarian cancer.