

## METABRIC Data Dictionary

### METABRIC Breast Cancer Cohort – Data Dictionary (Schema Format)

Each variable includes: Variable Name, Description, Data Type, Allowed Values / Range, and Notes.

#### Patient Demographics

Variable Name	Description	Data Type	Allowed Values / Range	Notes
Menopausal_Status	Menopausal state at diagnosis	Categorical	Pre, Post	Used in hormone therapy stratification
Age_at_Diagnosis	Age in years at diagnosis	Integer	20–90+	Median: 62
Age_Group	Age bracket for stratified survival	Categorical	<40, 40–49, 50–59, 60–69, 70+	Used in survival curve segmentation

#### Tumor Characteristics

Variable Name	Description	Data Type	Allowed Values / Range	Notes
PAM50_Subtype	Molecular subtype	Categorical	LumA, LumB, Her2, Basal, Claudin-low, Normal	HER2 status varies by subtype
HER2_Status	HER2 receptor status	Binary	Positive, Negative	Key for treatment eligibility
Histologic_Grade	Tumor grade	Ordinal	1, 2, 3	Higher grade = worse prognosis
Tumor_Stage	Clinical tumor stage	Ordinal	0–4	Used in recurrence analysis
Tumor_Cellularity	Cell density	Categorical	Low, Moderate, High	Correlates with histologic grade
Cancer_Type_Detailed	Histopathological diagnosis	Categorical	Invasive Ductal, Lobular, Mixed, Mucinous	ICD subtype granularity

 **Integrative Clusters**

Variable Name	Description	Data Type	Allowed Values / Range	Notes
Integrative_Cluster	Genomic cluster assignment	Categorical	1–10, 4ER+, 4ER-	Used in molecular stratification

 **Treatment Variables**

Variable Name	Description	Data Type	Allowed Values / Range	Notes
Hormone_Therapy	Hormone therapy administered	Binary	Yes, No	Stratified by menopausal status
Breast_Surgery_Type	Type of surgery	Categorical	Mastectomy, Breast Conserving	Used in treatment pattern analysis
Chemotherapy	Chemotherapy administered	Binary	Yes, No	Optional field if available

 **Outcomes**

Variable Name	Description	Data Type	Allowed Values / Range	Notes
Overall_Survival_Months	Survival duration	Integer	0–240+	Median: 116 months
Survival_Status	Vital status	Binary	Living, Deceased	42.7% living, 57.3% deceased
Recurrence_Status	Cancer recurrence	Binary	Recurred, Not Recurred	Stratified by tumor stage and grade
Percent_Recurred	Recurrence rate by subgroup	Float (%)	0–100	Used in grade-based outcome analysis

 **Cohort Summary Metrics (Metadata)**

Variable Name	Description	Data Type	Allowed Values / Range	Notes
Total_Patients	Total cohort size	Integer	1917	Static
Living_Patients	Number alive	Integer	818	Static
Median_Overall_Survival	Median survival	Integer	116	Static

---