

Data Dictionary Name: Pathways Cancer Registry

Dataset Name: df_cancerregistry_20240116.sas7bdat; df_cancerregistry_20240116.xlsx

Description: The Pathways cancer registry table contains tumor, patient, and treatment characteristics for all Pathways participants. These data were procured from the Kaiser Permanente Northern California Cancer Registry dataset which is periodically reported to the California Cancer Registry and the Surveillance, Epidemiology, and End-Results (SEER) Program of the National Cancer Institute (NCI). In some cases, certain variables are renamed, reformatted, or supplemented with additional data from other sources in order to accommodate the needs of the study. In such cases, we have described the nature of these changes in the notes section below.

Key Variables:

*Study identification number (studyid)

Cancer registry diagnosis date (cr_dx_date)

AJCC stage (pw_ajcc_stg)

Chemotherapy status (pw_chemo_yn)

Radiation status (pw_rad_yn)

Hormone therapy status (pw_horm_yn)

**primary key (use this variable to link to other datasets)*

Variable	Description	Values	Notes	Format Name
studyid	PATHWAYS PARTICIPANT STUDY IDENTIFICATION NUMBER			
cr_behav	TUMOR BEHAVIOR	0 = BENIGN 1 = BORDERLINE 2 = IN-SITU 3 = MALIGNANT		CR_BEHAVF
cr_dx_date	BREAST CANCER KPNC CANCER REGISTRY DIAGNOSIS DATE			DATE9
cr_fu_date	DATE OF LAST PATIENT FOLLOW-UP (DATE OF LAST CONTACT IF LIVING, OR DATE OF DEATH IF DECEASED)			DATE9
cr_icdosite	INTERNATIONAL CLASSIFICATION OF DISEASES FOR ONCOLOGY, THIRD EDITION (ICD-O-3) SURVEILLANCE, EPIDEMIOLOGY, AND END RESULTS (SEER) TUMOR SITE	C500 = NIPPLE C501 = CENTRAL PORTION OF BREAST C502 = UPPER- INNER QUADRANT OF BREAST		PW_ICDOSITEF

Variable	Description	Values	Notes	Format Name
		C503 = LOWER- INNER QUADRANT OF BREAST C504 = UPPER- OUTER QUADRANT OF BREAST C505 = LOWER- OUTER QUADRANT OF BREAST C506 = AXILLARY TAIL OF BREAST C508 = OVERLAPPING LESION OF BREAST C509 = BREAST, NOS		
cr_immuno_date	DATE OF FIRST IMMUNOTHERAPY TREATMENT			DATE9

Variable	Description	Values	Notes	Format Name
cr_morph	INTERNATIONAL CLASSIFICATION OF DISEASES FOR ONCOLOGY, THIRD EDITION (ICD-O-3) SURVEILLANCE, EPIDEMIOLOGY, AND END RESULTS (SEER) COMMON HISTOLOGIC/MORPHOLOGIC GROUP			
cr_ncd_date	DIAGNOSTIC PROCEDURE OR BIOPSY DATE			DATE9
cr_oth_date	DATE OF FIRST CANCER TREATMENT CLASSIFIED AS OTHER			DATE9
cr_surg_date	EARLIEST SURGERY DATE (EXCLUDING BIOPSY)			DATE9
enrolled	PATHWAYS ENROLLMENT STATUS	0 = NOT ENROLLED 1 = ENROLLED		PW_ENROLLEDF
pw_ajcc_stg	AMERICAN JOINT COMMITTEE OF CANCER (AJCC) STAGE (AJCC CANCER STAGING MANUAL, 6TH AND 7TH EDITIONS)	00-02 = 0 10-24 = I 30-43 = II 50-63 = III 70-74 = IV 88 = N/A	AJCC Cancer Staging Manual 6th Edition was used for breast cancer diagnoses between 2004-2009. AJCC Cancer Staging Manual 7th Edition was used for breast	PW_AJCCSTGF

Variable	Description	Values	Notes	Format Name
		90 = OCCULT 99 = UNKNOWN	cancer diagnoses between 2010-2017.	
pw_chemo_date	DATE OF FIRST CHEMOTHERAPY TREATMENT AFTER CANCER DIAGNOSIS			DATE9
pw_chemo_yn	CHEMOTHERAPY TREATMENT	0 = NO 1 = YES 9 = UNKNOWN	This variable was created utilizing data sources according to the following hierarchy: (1) KPNC Cancer Registry data and (2) KPNC Infusion data.	PW_YESNOF
pw_cs_sz	COLLABORATIVE STAGE (CS) TUMOR SIZE	0 = NO MASS/TUMOR FOUND 1 - 988 = EXACT SIZE IN MILLIMETERS 989 = 989 MM OR LARGER 990 = MICROINVASION, LESS THAN 1 MM, T1MI NO OTHER INFO		PW_CSSZF

Variable	Description	Values	Notes	Format Name
		991 = LESS THAN 1 CM, T1B NO OTHER INFO 992 = LESS THAN 2 CM/GREATER THAN 1 CM/BETWEEN 1 AND 2 CM, T1 [NOS] OR T1C [NOS] NO OTHER INFO 993 = LESS THAN 3 CM/GREATER THAN 2 CM/BETWEEN 2 AND 3 CM 994 = LESS THAN 4 CM/GREATER THAN 3 CM/BETWEEN 3 AND 4 CM 995 = LESS THAN 5 CM/GREATER THAN 4 CM/BETWEEN 4 AND 5 CM, T2 NO OTHER INFO 996 = MAMMOGRAPHIC OR XEROGRAPHIC DX ONLY, NO SIZE GIVEN, CLINICALLY NOT PALPABLE		

Variable	Description	Values	Notes	Format Name
		997 = PAGET DISEASE OF NIPPLE WITH NO DEMONSTRABLE TUMOR 998 = DIFFUSE 999 = UNKNOWN SIZE NOT STATED, SIZE CANNOT BE ASSESSED, NOT DOCUMENTED IN PT RECORD		
pw_csln_class	COLLABORATIVE STAGE (CS) REGIONAL LYMPH NODE STAGING	1 = NONE 2 = N1 3 = N2 OR N3 4 = UNKNOWN	CS codes 000 and 050 considered "None". CS codes 130, 150, 155, 250, 257, 258, 260, 520, 600, 610, 620, 630, 710, 720, 730, 735 and 740 considered "N1". CS codes 280, 290, 300, 500, 510, 735, 745, 748, 750, 755, 760, 763, 764, 765, 768, 770, 780, 790, 800, 805, 810, 815 and 820 considered "N2 or N3".	PW_CSLNCLASSF

Variable	Description	Values	Notes	Format Name
pw_dxdt	BREAST CANCER DIAGNOSIS DATE USED BY PATHWAYS STUDY		This diagnosis date was identified in the KPNC cancer registry and then pathologically reviewed and confirmed by Roswell Park Comprehensive Cancer Center.	DATE9
pw_er	ESTROGEN RECEPTOR (ER) STATUS	0 = NOT DONE 1 = POSITIVE 2 = NEGATIVE 3 = BORDERLINE 8 = ORDERED, N/A 9 = UNKNOWN . = MISSING	Estrogen receptor status was identified by manual medical chart review. If the status was not available in the chart, the Collaborative Stage (CS) Site-Specific Factor 1 was utilized. SSF1 code 10 considered positive. SSF1 code 20 considered negative. SSF1 code 30 considered borderline.	PW_TMARKERF
pw_grade	TUMOR GRADE	1 = GRADE I - WELL DIFFERENTIATED 2 = GRADE II - MODERATELY DIFFERENTIATED		PW_GRADEF

Variable	Description	Values	Notes	Format Name
		3 = GRADE III - POORLY DIFFERENTIATED 4 = GRADE IV - UNDIFFERENTIATED 5 = T-CELL (IF APPLICABLE) 6 = B-CELL (IF APPLICABLE) 7 = NULL-CELL (IF APPLICABLE) 8 = NK-CELL (IF APPLICABLE) 9 = GRADE OR DIFFERENTIATION NOT DETERMINED		
pw_her2	HER2/NEU PROTEIN STATUS	0 = NOT DONE 1 = POSITIVE 2 = NEGATIVE 3 = BORDERLINE 8 = ORDERED, N/A 9 = UNKNOWN	This variable was created utilizing data sources according to the following hierarchy: (1) manual medical chart reviewed data, (2) Fluorescence In Situ Hybridization (FISH) test data, (3)	PW_TMARKERF

Variable	Description	Values	Notes	Format Name
		. = MISSING	immunohistochemistry (IHC) data, (4) Collaborative Stage (CS) Site-Specific Factor 15, and (5) KPNC Cancer Registry data. SSF15 code 10 considered positive. SSF15 code 20 considered negative. SSF15 code 30 considered borderline.	
pw_herceptin_1stdate	DATE OF FIRST HERCEPTIN TREATMENT			DATE9
pw_herceptin_yn	HERCEPTIN TREATMENT	0 = NO 1 = YES 9 = UNKNOWN		PW_YESNOF
pw_horm_date	DATE OF FIRST HORMONE THERAPY TREATMENT AFTER CANCER DIAGNOSIS			DATE9

Variable	Description	Values	Notes	Format Name
pw_horm_yn	HORMONE THERAPY TREATMENT	0 = NO 1 = YES 9 = UNKNOWN	This variable was created utilizing data sources according to the following hierarchy: (1) KPNC cancer registry data and (2) KPNC prescriptions data.	PW_YESNOF
pw_laterality	TUMOR LATERALITY	0 = NOT A PAIRED SITE 1 = RIGHT 2 = LEFT 3 = UNILATERAL, RIGHT OR LEFT SIDE, ORIGIN UNKNOWN 4 = BILATERAL, ORIGIN UNKNOWN 5 = PAIRED SITE, MIDLINE TUMOR 9 = PAIRED SITE, BUT NO INFORMATION REGARDING LATERALITY		PW_LATF

Variable	Description	Values	Notes	Format Name
pw_met_site_1	FIRST METASTASIS SITE		Identified by manual medical chart review.	
pw_met_site_2	SECOND METASTASIS SITE		Identified by manual medical chart review.	
pw_met_site_3	THIRD METASTASIS SITE		Identified by manual medical chart review.	
pw_met_site_4	FOURTH METASTASIS SITE		Identified by manual medical chart review.	
pw_metas	COLLABORATIVE STAGE (CS) METASTASES AT DIAGNOSIS	0 = NO 1 = YES 9 = UNKNOWN	CS codes 00 considered "no". CS codes 10, 40, 42, 44 and 50 considered and 050 considered "yes".	PW_YESNOF
pw_multiple_tumors	SINGLE OR MULTIPLE TUMORS DIAGNOSED ON THE SAME PATHWAYS CANCER REGISTRY DIAGNOSIS DATE	1 = QUAL TUMOR UNIQUE 2 = MULT TUMORS BILATERAL 3 = MULT TUMORS UNILATERAL		PW_MULTIPLE_TUMORSF
pw_newprim_date	DATE OF FIRST NEW PRIMARY TUMOR POST INITIAL BREAST CANCER DIAGNOSIS			DATE9

Variable	Description	Values	Notes	Format Name
pw_nodal	COLLABORATIVE STAGE (CS) REGIONAL LYMPH NODES INVOLVEMENT	0 = NOT DONE 1 = POSITIVE 2 = NEGATIVE 3 = BORDERLINE 8 = ORDERED, N/A 9 = UNKNOWN . = MISSING	CS code 000 considered negative. CS codes 050-820 considered positive.	PW_TMARKERF
pw_posnodes	COLLABORATIVE STAGE (CS) NUMBER OF PATHOLOGICALLY POSITIVE REGIONAL LYMPH NODES	00 = ALL NEGATIVE 01 - 89 = NUMBER OF NODES POSITIVE 90 = 90+ POSITIVE NODES 95 = POSITIVE ASPIRATION 97 = POSITIVE NODES - NUMBER UNSPECIFIED 98 = NONE EXAMINED 99 = UNKNOWN		PW_POSNODESF

Variable	Description	Values	Notes	Format Name
pw_pr	PROGESTERONE RECEPTOR (PR) STATUS	0 = NOT DONE 1 = POSITIVE 2 = NEGATIVE 3 = BORDERLINE 8 = ORDERED, N/A 9 = UNKNOWN . = MISSING	Progesterone receptor status was identified by manual medical chart review. If the status was not available in the chart, the Collaborative Stage (CS) Site-Specific Factor 2 was utilized. SSF2 code 10 considered positive. SSF2 code 20 considered negative. SSF2 code 30 considered borderline.	PW_TMARKERF
pw_prior_tumor	TUMOR PRIOR TO PATHWAYS QUALIFYING CANCER DIAGNOSIS	0 = NO 1 = YES 9 = UNKNOWN		PW_YESNOF
pw_prior_tumor_type1	TUMOR TYPE (OF FIRST TUMOR) PRIOR TO PATHWAYS QUALIFYING CANCER DIAGNOSIS	1 = NONE 2 = DUCTAL CARCINOMA IN SITU (DCIS) 3 = NON-INVASIVE TUMOR 4 = OTHER INVASIVE TUMOR	If a person had a tumor prior to their Pathways qualifying cancer diagnosis, this variable specifies the type of tumor.	PW_PRIOR_TUMORF

Variable	Description	Values	Notes	Format Name
pw_prior_tumor_type2	TUMOR TYPE (OF SECOND TUMOR) PRIOR TO PATHWAYS QUALIFYING CANCER DIAGNOSIS	1 = NONE 2 = DUCTAL CARCINOMA IN SITU (DCIS) 3 = NON-INVASIVE TUMOR 4 = OTHER INVASIVE TUMOR	If a person had two tumors prior to their Pathways qualifying cancer diagnosis, this variable specifies the type of the 2nd tumor.	PW_PRIOR_TUMORF
pw_prior_tumor_type3	TUMOR TYPE (OF THIRD TUMOR) PRIOR TO PATHWAYS QUALIFYING CANCER DIAGNOSIS	1 = NONE 2 = DUCTAL CARCINOMA IN SITU (DCIS) 3 = NON-INVASIVE TUMOR 4 = OTHER INVASIVE TUMOR	If a person had three tumors prior to their Pathways qualifying cancer diagnosis, this variable specifies the type of the 3rd tumor.	PW_PRIOR_TUMORF
pw_rad_date	DATE OF FIRST RADIATION TREATMENT AFTER CANCER DIAGNOSIS			DATE9

Variable	Description	Values	Notes	Format Name
pw_rad_yn	RADIATION TREATMENT	0 = NO 1 = YES 9 = UNKNOWN	This variable was created utilizing data sources according to the following hierarchy: (1) KPNC cancer registry data and (2) KPNC procedures data.	PW_YESNOF
pw_recur_date	FIRST DATE OF RECURRENCE AFTER DISEASE-FREE PERIOD			DATE9
pw_subtype_gallen	ST GALLEN MOLECULAR SUBTYPE	1 = LUMINAL A 2 = LUMINAL B 3 = HER2E 4 = NNN		PW_GALLENF
pw_subtype_ihc	IMMUNOHISTOCHEMICAL MOLECULAR SUBTYPE	1 = LUMINAL A 2 = LUMINAL B 3 = NNN 4 = HER2OVER		PW_IHCF
pw_surg_sum	SUMMARY OF SURGICAL PROCEDURE	0 = NO SURGERY 10 - 14 = LOCAL TUMOR DESTRUCTION, NOS		PW_SURGF

Variable	Description	Values	Notes	Format Name
		20 - 29 = PARTIAL MASTECTOMY 30 = SUBCUTANEOUS MASTECTOMY 40 - 49 = SIMPLE MASTECTOMY 50 - 59 = MODIFIED RADICAL MASTECTOMY 60 - 69 = RADICAL MASTECTOMY, NOS 70 - 79 = EXTENDED RADICAL MASTECTOMY 80 - 89 = MASTECTOMY, NOS 90 = SURGERY, NOS 98 = MISCODE 99 = UNKNOWN		
pw_taxane_1stdate	DATE OF FIRST TAXANE TREATMENT			DATE9

Variable	Description	Values	Notes	Format Name
pw_taxane_yn	TAXANES TREATMENT	0 = NO 1 = YES 9 = UNKNOWN		PW_YESNOF

This data dictionary was created on January 16, 2024.