FRAMINGHAM HEART STUDY CODING MANUAL

Cancer Validated by Medical Records

COHORT: Original, Offspring, Generation 3, New Offspring Spouse, Omni 1, Omni 2

TIME INTERVAL: Cancer Cases Reviewed through February 2019 for Offspring and Generation 3 cohorts, through September 2013 for all other cohorts

DATASET NAME: vr_cancer_2019_a_1162s

OBSERVATIONS: 6362

Variables that define a unique record: SHAREID, TOPO, LOCATION, HIST, BEHAVIOR, GRADE.

This dataset has 6,362 cancer cases among 4,777 individuals.

All known cancer cases were validated using medical records (pathology reports). Occasionally a pathology report was unavailable. In these cases, other confirming evidence was used (see DIAG_TYP variable). This dataset contains only primary cancers (ICD-O behavior codes of 2 or 3). This dataset does not include metastatic cancers (ICD-O behavior code of 6). Only first occurrence of site and cell type is included in this dataset.

Most cancer records were reviewed using the FHS version of the 1976 ICD-O-1 code book. Refer to ICD-O-3 for details of morphology codes of lymphomas and several other morphologies.

To exclude non-melanotic skin cancers from this dataset:
Delete cases where TOPO = 173 and (HIST < 8720 or HIST > 8790) but keep HIST=8247.

This dataset replaces vr_cancer_2013_a_0018. Note that the 2019 cancer dataset updated cancer cases for Offspring and Generation 3 cohorts only.

This dataset is a modified version of the original dataset. Some observations were deleted due to consent restrictions. This coding manual is based on the original dataset. Therefore the value ranges, number of observations and/or cohorts found in the modified dataset, may differ from those stated in this manual.

IDTYPE

Variable name: IDTYPE

Description: Framingham Heart Study cohort identifier

Code or Value	Description	Count
0	Original Cohort	2446
1	Offspring Cohort	3162
2	New Offspring Spouse Cohort	43
3	Generation 3 Cohort	622
7	Omni 1 Cohort	68
72	Omni 2 Cohort	21

Note: IDTYPE and ID taken together form a unique study participant identification number.

ID

Variable name: ID

Description: Framingham Heart Study participant identifier within a cohort

Note: IDTYPE and ID taken together form a unique study participant identification number.

ID deleted - use SHAREID

SHAREID

Variable name: SHAREID

Description: Unique study participant identification number

D_DATE

Variable name: D_DATE
Description: Diagnosis date of tumor or cancer
Units: Number of days since Exam 1

Code or Value	Description	Count
-10345 — 22506	Range of values	6362

Note: Date was taken from pathology reports if available. If the date was taken from other records because the pathology report was not available or if diagnosis was made before the biopsy then the variable EXACT = 1.

If month and year were available, but not day, the 15th was used as the day, and the variable EXACT = 1. If only the year was available, JULY 4 was used as the month and day, and the variable EXACT = 1.

In some cases if biopsy was done at autopsy, then D_DATE was changed to precede DATE OF DEATH by one day.

D_DATE includes month, day and year, however this coding manual displays only month and year to preserve confidentiality.

EXACT

Variable name: EXACT
Description: Type of Date - diagnosis date qualifier - indicates the quality of D_DATE

Code or Value	Description	Count
0	Month, date and year of diagnosis date taken from pathology report or other test report.	5959
1	Estimated diagnosis date as described in variable D_DATE	403

DIAG_TYP

Variable name: DIAG_TYP
Description: Method of diagnosis - represents the basis of diagnosis

Code or Value	Description	Count
1	Microscopically confirmed	6110
2	Clinically diagnosed	211
3	Death certificate only	41

TOPO

Variable name: TOPO

Description: Topography - describes site - corresponds to first three digits of topography code WHO ICD-O

Code or Value	Description	Count
140	Lip	0
141	Tongue	15
142	Major salivary gland	7
144	Floor of mouth/Gum	10
145	Other and unspecified parts of mouth	18
146	Oropharynx	16
148	Hypopharynx	10
149	Pharynx, Nasopharynx and ill-defined sites in lip, oral cavity and pharynx	8
150	Esophagus	35
151	Stomach	106
152	Small intestine	17
153	Colon	391
154	Rectum, rectosigmoid junction, anal canal and anus, not otherwise specified	155
155	Liver and intrahepatic bile ducts	28
156	Gallbladder and extrahepatic bile ducts	31
157	Pancreas	102
158	Retroperitoneum and peritoneum	15
160	Nasal cavities, accessory sinuses, middle ear and inner ear	7
161	Larynx	41
162	Trachea bronchus, and lung	513
163	Pleura	13
169	Hematopoietic and reticuloendothelial systems	241
170	Bones, joints and articular cartilage	9
171	Connective, subcutaneous and other soft tissues	22
173	Skin (excludes skin of labia majora, skin of vulva, skin of penis and skin of scrotum)	2236
174	Female breast (excludes skin of breast)	752

175	Male breast (excludes skin of breast)	10
179	Uterus, not otherwise specified	5
180	Cervix uteri	76
181	Placenta	0
182	Corpus uteri or endometrium	119
183	Ovary, fallopian tube and broad ligament	70
184	Other and unspecified female genital organs	20
185	Prostate gland	551
186	Testis	21
187	Penis and other male genital organs	6
188	Urinary bladder	243
189	Kidney and other urinary organs	105
190	Eye and lacrimal gland	10
191	Brain	61
193	Thyroid gland	60
195	Other ill-defined sites; see NOTE	12
196	Lymph nodes	112
199	Unknown primary site	83

Note: TOPO modified - cancer site categories with less than five cases were recategorized under "Other ill-defined sites" to preserve participant confidentiality. The following are the cancer site categories that were recategorized:
159 = Other and ill-defined sites within digestive organs and peritoneum
164 = Thymus, heart and mediastinum
165 = Other and ill-defined sites within respiratory system and intrathoracic organs
192 = Other and unspecified parts of the nervous system (excludes peripheral nerves, sympathetic and parasympathetic nerves and ganglia)
194 = Other endocrine glands

LOCATION

Variable name: LOCATION
Description: Location - describes subdivision of site - corresponds to fourth digit of topography code WHO ICD-O

Code or Value	Description	Count
0 — 9	Range of values	6362

HIST

Variable name:	HIST
Description:	Histology or morphology - describes cell histopathology - corresponds to four digit morphology code WHO ICD-O

Code or Value	Description	Count
8000 — 9990	Range of values	6362

Note: Additional morphology codes from the ICD-O-3 code book have been added to the FHS cancer coding system:

8046 - Non-small cell carcinoma

8077 - Squamous intraepithelial neoplasia,
      Cervical intraepithelial neoplasia,
      Vaginal intraepithelial neoplasia,
      Vulvar intraepithelial neoplasia,
      Anal intraepithelial neoplasia

8130 - Papillary transitional cell carcinoma

8148 - Glandular intraepithelial neoplasia,
      Prostatic intraepithelial neoplasia

8162 - Klatskin tumor

8253 - Bronchiolo-alveolar carcinoma, mucinous

8255 - Adenocarcinoma with mixed subtypes

8317 - Renal cell carcinoma, chromophobe type

8522 - Intraductal carcinoma and lobular carcinoma in situ

8858 - Dedifferentiated liposarcoma

9590-9769 - Hodgkin's and non-Hodgkin lymphomas

9590-9596 - Malignant lymphomas, NOS or diffuse

9650-9667 - Hodgkin's lymphoma

9670-9729 - Non-Hodgkin's lymphomas

9670-9699 - Mature B-cell lymphomas
      9673 - Mantle cell lymphoma
      9680 - Malignant lymphoma, large B-cell, diffuse or malignant
          lymphoma, large cell

9700-9719 - Mature T-cell and NK-cell lymphomas

9727-9729 - Precursor cell lymohoblastic lymphomas

9731-9734 - Plasma cell tumors

9740-9742 - Mast cell tumors

9750-9758 - Neoplasms of histiocytes and accessory lymphoid cells

9760-9769 - Immunoproliferative diseases

9872 - Acute myeloid leukemia, minimal differentiation

9895 - Acute myeloid leukemia with multilineage dysplasia

9987 - Therapy related myelodysplastic syndrome

9989 - Myelodysplastic syndrome, NOS

BEHAVIOR

Variable name: BEHAVIOR

Description: Behavior - decribes the degree of malignancy - corresponds to the fifth digit behavior code WHO ICD-O

Code or Value	Description	Count
0	Benign	0
1	Uncertain, borderline	0
2	Carcinoma in situ, non-invasive	574
3	Malignant, primary site	5788
6	Malignant, metastatic site	0
9	Malignant, uncertain whether primary or metastatic site	0

Note: This dataset contains only cases where BEHAVIOR = 2 or BEHAVIOR = 3. Metastatic code (BEHAVIOR = 6) was not used; rather the location of the primary site was used. If the primary site is unknown, the code 199 is used for the site.

GRADE

Variable name: GRADE

Description: Grade - correspnds to code for histologic grading and differentiation WHO ICD-O

Code or Value	Description	Count
1	Grade I - well-differentiated, differentiated	627
2	Grade II - moderately well-differentiated, moderately differentiated	826
3	Grade III - poorly differentiated	1005
4	Grade IV - undifferentiated, anaplastic	240
9	Grade or differentiation not determined, stated, or applicable	3664

ESTRGN

Variable name: ESTRGN
Description: Estrogen receptors

Code or Value	Description	Count
0	Negative	59
1	Weakly positive	10
2	Positive	296
8	Not done	136
.	Unknown or collected on a form that did not contain this variable	5861

Note: This variable was added to the data collection form in version 10/2007.

PRGSTRN

Variable name: PRGSTRN
Description: Progesterone receptors

Code or Value	Description	Count
0	Negative	87
1	Weakly positive	25
2	Positive	239
8	Not done	148
.	Unknown or collected on a form that did not contain this variable	5863

Note: This variable was added to the data collection form in version 10/2007.

H2NI

Variable name: H2NI

Description: HER2NEU (human epidermal growth factor receptor 2) by IHC (immunohistochemistry)

Code or Value	Description	Count
0	Negative, 0 or 1+	177
1	Weakly positive, 2+	20
2	Positive, 3+	24
8	Not done	339
.	Unknown or collected on a form that did not contain this variable	5802

Note: This variable was added to the data collection form in version 10/2007.

H2NF

Variable name: H2NF

Description: HER2NEU (human epidermal growth factor receptor 2) by FISH (fluorescence in situ hybridization)

Code or Value	Description	Count
0	Unamplified	60
1	Amplified by FISH	8
8	Not done	482
.	Unknown or collected on a form that did not contain this variable	5812

Note: This variable was added to the data collection form in version 10/2007.

TISSUE_1

Variable name: TISSUE_1

Description: Tissue #1 related to Gleason score for prostate cancer

Code or Value	Description	Count
1 — 5	Range of values	340
.	Unknown or collected on a form that did not contain this variable	6022

Note: Classifying prostate cancer cells in terms of how abnormal they appear when examined under a microscope. Grades range from 1 to 5 with 1 being the least aggressive and 5 the most aggressive.

This variable was added to the data collection form in version 10/2007.

TISSUE_2

Variable name: TISSUE_2
Description: Tissue #2 related to Gleason score for prostate cancer

Code or Value	Description	Count
1 — 5	Range of values	328
.	Unknown or collected on a form that did not contain this variable	6034

Note: Classifying prostate cancer cells in terms of how abnormal they appear when examined under a microscope. Grades range from 1 to 5 with 1 being the least aggressive and 5 the most aggressive.

This variable was added to the data collection form in version 10/2007.

GLEASON

Variable name: GLEASON
Description: Gleason score: classifying prostate cancer cells

Code or Value	Description	Count
2 — 10	Range of values	364
.	Unknown or collected on a form that did not contain this variable	5998

Note: Gleason score (total between 2-10) TISSUE_1 + TISSUE_2 = GLEASON
If only GLEASON is given, TISSUE_1 and TISSUE_2 = "."
If no scores are given, GLEASON = "." and TISSUE_1 and TISSUE_2 = "."

This variable was added to the data collection form in version 10/2007.

INDICATOR

Variable name: INDICATOR
Description: Indicates ICD-O code book used for coding

Code or Value	Description	Count
0	ICD-O-1 code book	5972
1	ICD-O-3 code book	355
2	ICD-O-3 code book using 'Coding Guidelines for Topography and Morphology' in front of code book	35

VERSION

Variable name: VERSION
Description: Version date of the data collection form

Code or Value	Description	Count
10/2007	Data collection form revised on 10/2007	559
03/2012	Data collection form revised on 03/2012	1083
.	Unknown	4720

Note: Prior to 10/2007 data collection forms did not have version dates.

VERSION deleted - for administrative purposes only