Medicines & Healthcare products Regulatory Agency

CPRD

Cancer Registration Data Dictionary

Version 10.2

Date: 3 February 2023

NIHR | National Institute for Health Research

Documentation Control Sheet

Over time, it may be necessary to issue amendments or clarifications to parts of this document. This form must be updated whenever changes are made.

Version	Summary of Change	Prepared By	Reviewed By
7.0	Initial draft	Helen Strongman	Rachael Williams
7.1	Modified	Helen Strongman	Eleanor Yelland
8.0	Modified	Rachael Williams	Eleanor Yelland
8.1	Modified	Eleanor Yelland	Helen Booth
9.0	Modified	Eleanor Yelland	Sonia Coton
10.0	Modified	Eleanor Yelland	Hilary Shepherd
10.1	Modified	Hilary Shepherd	Eleanor Yelland/Susan Hodgson
10.2	Modified	Jessie Oyinlola	Hilary Shepherd

Version 7.0

- Cancer Registration Data dictionary separated from NCRAS documentation. No changes to data structure between set 13 (up to 2014) and set 14 (up to 2015)
- Created separate data documentation and data dictionary files

Version 7.1

- Added tumour identifier to patient file for clarity (this has been available since set 14)

Version 8.0

- Refreshed for set 16

Version 8.1

- Refreshed for set 17
- Minor updates to variable descriptions

Version 9.0

- Updated for set 18

Version 10.0

- Updated for set 19
- Addition of Charlson comorbidity variables
- Minor updates to variable names and re-ordering of variables
- Additional data minimisation variables available
- Addition of code description variables

Version 10.1

- Updated for set 21
- Added variables from v4.3 of the ODR
- Renamed tables in line with ODR, added vitalstatus, vitaldate, creg_name
- Added DOIs

Version 10.2

- Updated list of available variables and descriptions. Removed sex, birthyear, age and fiveyearageband as routinely available as they are available in CPRD primary care data. Removed variables relating to month and year as they are derived from full date variables (diagnosisdatebestmonth, diagnosisdatebestyear from Tumour table, and eventmonth, eventyear from both Tumour and Treatment tables)

- Removed set specific information

DOI
Please cite in any publications using these data:

CPRD GOLD Cancer Registration August 2021 - https://doi.org/10.48329/541y-nh70

CPRD Aurum Cancer Registration August 2021 - https://doi.org/10.48329/5sm7-3209

1. Patient demographics

Column description	*Column name*	*Details*	*Field Type*	*Valid Content*
CPRD patient Identifier	e_patid	Unique patient identifier based on CPRD primary care data – pseudonymised. In some cases, the same person may have multiple patient IDs.	NUMBER	Number
CR patient Identifier	e_cr_patid	Unique patient identifier based on NCRAS data patient identifier – pseudonymised. In some cases, the same person may have multiple patient IDs. Patient IDs will be retained even after two patient records are found to be the same person.	NUMBER	Number
Ethnicity code	ethnicity	The 16+1 ethnic data categories defined in the 2001 census is the national mandatory standard for the collection and analysis of ethnicity.	TEXT	A = (White) British, B =(White) Irish, C = Any other White background, D = White and Black Caribbean, E = White and Black African, F = White and Asian, G = Any other mixed background, H = Indian, J = Pakistani, K = Bangladeshi, L = Any other Asian background, M = Caribbean, N = African, P = Any other Black background, R = Chinese, S = Any other ethnic group, Z = Not stated, X = Not Known.
Ethnicity name	ethnicityname	Description of the ethnicity variable	TEXT	

2. Tumour diagnosis details

Column description	*Column name*	*Details*	*Field Type*	*Valid Content*
CPRD patient Identifier	e_patid	Unique patient identifier based on CPRD primary care data – pseudonymised. In some cases, the same person may have multiple patient IDs.	NUMBER	Number
CR patient Identifier	e_cr_patid	Unique patient identifier based on NCRAS data – pseudonymised. In some cases, the same person may have multiple patient IDs. Patient IDs will be retained even after two patient records are found to be the same person.	NUMBER	Number
CR tumour identifier	e_cr_id	Unique tumour identifier based on NCRAS data – pseudonymised. As someone can have more than one tumour a patient can have multiple unique tumour IDs.	NUMBER	Number
Diagnosis date	diagnosisdatebest	Diagnosis date of the patient, as defined by the UKIACR, per the recommendations from the ENCR. Month and year diagnosis variables are derived from this field.	DATE	dd/mm/yyyy
Date of diagnosis check flag	diagnosisdateflag	Should be used in combination with diagnosisdatebest, should the project require an understanding if the date is imputed. Imputation of dates follows rules agreed by UKACR DQAR sub-group (August 2010). Blank field indicates that date imputation did not occur.	NUMBER	0 = Date fully specified, 1 = Month and year specified, 2 = Year specified, 3 = 3 = Date less specific than any of above.
Basis of diagnosis of the tumour	basisofdiagnosis	Basis of diagnosis of the tumour according to all the data received by the registry.	NUMBER	Non-microscopic: 0 = Death certificate. 1 = Clinical: Diagnosis made before death without (2-7). 2 = Clinical investigation: Includes all diagnostic

				techniques without a tissue diagnosis. 4 = Specific tumour markers: Includes biochemical and/or immunological markers which are site specific. Microscopic: 5 = Cytology: Examination of cells whether from a primary or secondary site, including fluids aspirated using endoscopes or needles. Also including microscopic examination of peripheral blood films and trephine bone marrow aspirates. 6 = Histology of a metastases: Includes autopsy specimens. 7 = Histology of a primary tumour: Includes all cutting and bone marrow biopsies. Also includes autopsy specimens of a primary tumour. 9 = Unknown, e.g. PAS or HISS record only.
Site of neoplasm	site_icd10_o2	Describes site of origin of the neoplasm.	TEXT	Valid 4 digit ICD-10 codes

(4-character ICD-10-O2 code)				in the range C00-D48 plus D76, E85, O01, Q85 or blank.
Site of neoplasm (3-character ICD-10-O2 code)	site_icd10_o2_3char	Describes site of origin of the neoplasm.	TEXT	Valid 3 digit ICD-10 codes in the range C00-D48 plus D76, E85, O01, Q85 or blank.
Morphology of the cancer, in the ICD-10-O2 system	morph_icd10_o2	Morphology of cancer mapped to ICD-10-O2.	NUMBER	Number 8000-9990 or blank
Behaviour of the cancer, in the ICD-10-O2 system	behaviour_icd10_o2	Behaviour of cancer mapped to ICD-10-O2	TEXT	0 = Benign, 1 = Uncertain, 2 = In Situ, 3 = Malignant, 5 = Micro-Invasive, 6 = Malignant, Metastatic / Secondary Site, 9 = Malignant, Uncertain whether primary or metastatic, X = Unknown / Inapplicable.
Site of the cancer	site_coded	Site code of the cancer, in the coding system that the tumour was originally coded in. The coding system it is coded in is specified in coding_system/_desc. This variable (or site_coded_3char) should be selected if data prior to 1995 are being requested. Site_coded, site_coded_desc, site_coded_3char, coding_system, coding_system_desc, morph_coded, behaviour_coded, behaviour_coded_desc, histology_coded and histology_coded_desc can be provided together to ensure correct interpretation.	TEXT	
Site coded description	site_coded_desc	Description of the site code of the cancer	TEXT	
3 digit version of	site_coded_3char	Three digit version of SITE_CODED. This variable (or	TEXT	

site_coded		site_coded) should be selected if data prior to 1995 are being requested. Site_coded, site_coded_desc, site_coded_3char, coding_system, coding_system_desc, morph_coded, behaviour_coded, behaviour_coded_desc, histology_coded and histology_coded_desc can be provided together to ensure correct interpretation.		
The coding system used to register the tumour	coding_system	The coding system used to register the tumour (e.g. ICD10/O-2). Site_coded, site_coded_desc, site_coded_3char, coding_system, coding_system_desc, morph_coded, behaviour_coded, behaviour_coded_desc, histology_coded and histology_coded_desc can be provided together to ensure correct interpretation.	NUMBER	1 = ICD-8, 2 = ICD-9, 3 = ICD-10/O-2, 4 = ICD-10/O-3, 5 = ICD-O-3, 6 = ICD-7, 7 = ICD-8pre1971, 8 = ICD-O-2, 9 = ICD-O, 10 = ICD-O-3 (2011), 11 = ICD-10rev4/O-2, 12 = MOTNAC, 14 = SNOMED/O(TCR), 15 = SNOMED/O-1, 16 = SNOMED/O-2, 17 = SNOMED/O-3.
Text description of site of the cancer	coding_system_desc	Description of the coding system. Site_coded, site_coded_desc, site_coded_3char, coding_system, coding_system_desc, morph_coded, behaviour_coded, behaviour_coded_desc, histology_coded and histology_coded_desc can be provided together to ensure correct interpretation.	TEXT	
Morphology	morph_coded	Morphology of the cancer, in the coding system that the tumour was originally coded in, describing the cell type of malignant disease determined before the start	TEXT	

		of treatment. The relevance of tumour morphology differs across tumour site. Site_coded, site_coded_desc, site_coded_3char, coding_system, coding_system_desc, morph_coded, behaviour_coded, behaviour_coded_desc, histology_coded and histology_coded_desc can be provided together to ensure correct interpretation.		
Numeric behaviour code	behaviour_coded	Behaviour of the cancer. Site_coded, site_coded_desc, site_coded_3char, coding_system, coding_system_desc, morph_coded, behaviour_coded, behaviour_coded_desc, histology_coded and histology_coded_desc can be provided together to ensure correct interpretation.	NUMBER	0 = Benign, 1 = Uncertain, 2 = In Situ, 3 = Malignant, 5 = Micro-Invasive, 6 = Malignant, Metastatic / Secondary Site, 9 = Malignant, Uncertain whether primary or metastatic, X = Unknown / Inapplicable.
Description of behaviour code	behaviour_coded_desc	Description of behaviour code. Site_coded, site_coded_desc, site_coded_3char, coding_system, coding_system_desc, morph_coded, behaviour_coded, behaviour_coded_desc, histology_coded and histology_coded_desc can be provided together to ensure correct interpretation.	TEXT	
Histology	histology_coded	Histology code. Site_coded, site_coded_desc, site_coded_3char, coding_system, coding_system_desc, morph_coded, behaviour_coded, behaviour_coded_desc, histology_coded and histology_coded_desc can be provided together to ensure correct interpretation.	TEXT	Histology code - combines the morphology and behaviour codes. Name for the histology type, for example clear cell meningioma.

Description of Histology Code	histology_coded_desc	Description for the coding system for histology. Site_coded, site_coded_desc, site_coded_3char, coding_system, coding_system_desc, morph_coded, behaviour_coded, behaviour_coded_desc, histology_coded and histology_coded_desc can be provided together to ensure correct interpretation.	TEXT	
Grade of tumour	grade	Records the grade of the tumour, for tumours that are graded on a simple numeric 1-3 or 1-4 scale. In tumours containing several areas of different grade, the grade of the predominant component is recorded.	TEXT	GX = Grade of differentiation is not appropriate or cannot be assessed, G0 = Borderline malignancy, G1 = Well differentiated, G2 = Moderately differentiated, G3 = Poorly differentiated, G4 = Undifferentiated / anaplastic, G5 = T Cell, G6 = B Cell, G7 = Null cell.
Size of the largest dimension of the tumour	tumoursize	Diameter of a tumour in mm, largest if more than one.	NUMBER	Number or blank
Number of nodes excised	nodesexcised	The number of local and regional nodes reported as being positive.	NUMBER	Number or blank
Number of nodes involved	nodesinvolved	The number of local and regional nodes examined.	NUMBER	Number or blank
Tumour count	tumourcount	Count of every tumour associated with this e_cr_patid	NUMBER	Number

Big tumour count	bigtumourcount	Count of every tumour associated with this e_cr_patid in range C00-97 excluding C44	NUMBER	Number
Route to diagnosis code	route_code	The code assigned to a route for the purpose of the algorithm. Note: available for cancers diagnosed in 2006-2014. See BJC publication.	TEXT	
Finalised route to diagnosis code	final_route	The published route with all dataset types accounted for. Note: available for cancers diagnosed in 2006-2014. See BJC publication.	TEXT	Options: DCO, Emergency Presentation, GP Referral, Screening, TWW, Other Outpatient, Inpatient Elective, Unknown
Best 'registry' stage at diagnosis of the tumour	stage_best	STAGE_BEST is an NCRAS derived field using a combination of best T, N and M in "t_best", "n_best" and "m_best". Includes Ann Arbor staging for lymphomas. NB: It is not guaranteed that data from the individual t_best, n_best and m_best variables have been combined into this variable, so they should be used in parallel. 2013 onwards: updated method that includes some site specific staging e.g. FIGO for gynaecological tumours; Ann Arbor for lymphomas; Rai and/or Binet for Chronic Lymphocytic Leukaemia (CLL), and; ISS for myeloma where available Pre-2013: original stage_best based on TNM staging information	TEXT	0, 0A, 0B, 0C, 0IS = Stage 0. 1, 1A, 1A1, 1A2, 1A3, 1AE, 1AES, 1AEX, 1AEXS, 1AS, 1AX, 1B, 1B1, 1B2, 1BE, 1BES, 1BEX, 1BS, 1BX, 1C, 1C1, 1C2, IC3, 1E, 1ES, 1EX, 1S, 1X, 1XS = Stage 1. 2, 2A, 2A1, 2A2, 2AE, 2AES, 2AEX, 2AS, 2AX, 2B, 2BE, 2BEX, 2BS, 2BX, 2C, 2E, 2ES, 2EX, 2S, 2X = Stage 2. 3, 3A, 3A1, 3A1i, 3A1ii, 3A2, 3AE, 3AES, 3AEX, 3AS, 3AX, 3B, 3BE, 3BES, 3BEX, 3BEXS, 3BS, 3BX, 3BXS, 3C, 3C1, 3C2, 3D, 3E, 3ES, 3EX, 3S, 3X, 3XS = Stage 3. 4, 4A, 4AE, 4AES, 4AEX,

				4AEXS, 4AS, 4AX, 4AXS, 4B, 4BE, 4BES, 4BEX, 4BEXS, 4BS, 4BX, 4BXS, 4BS, 4BX, 4BXS, 4C, 4E, 4ES, 4EX, 4EXS, 4S, 4X = Stage 4. A, B, C = RaiBinet stage 5 = Outdated or invalid code 6 = not stageable. ? = insufficient information. U = unstageable. X = not staged.
T stage (Best)	t_best	Local size of the tumour. T stage flagged by the registry as the best.	TEXT	UICC code
N stage (Best)	n_best	How far the tumour has grown through local tissues and distant spread of disease with involvement of lymph nodes. N stage flagged by the registry as the best.	TEXT	UICC code
M stage (Best)	m_best	Metastases to other organs. M stage flagged by the registry as the best.	TEXT	UICC code
Stage system (Best)	stage_best_system	Version of the TNM classification of malignant cancers used to stage the tumour for the best TNM values. Staging system associated with stage_best.	TEXT	5th, 6th, 7th, 8th, UICC5, UICC6, UICC7, UICC8 AJCC7, AJCC8, ENETS 2007, ANNARBOR, FIGO, ISS, RaiBinet, UNKNOWN

T stage (pre-treatment)	t_img	The UICC code which classifies the size and extent of the primary tumour before treatment.	TEXT	UICC code
N stage (pre-treatment)	n_img	The UICC code which classifies the absence or presence and extent of regional lymph node metastases before treatment	TEXT	UICC code +, 0, 1, 1a, 1b, 1c, 1mi, 2, 2a, 2b, 2c, 3, 3a, 3b, 3c, X
M stage (pre-treatment)	m_img	The UICC code which classifies the absence or presence of distant metastases pre-treatment	TEXT	0 = no distant metastasis 1, 1a, 1b, 1c, 1e = distant metastasis X = unknown
Stage at diagnosis (derived from imaging)	stage_img	Combination of imaging T, N and M in “t_img”, “n_img” and “m_img”. Includes Ann Arbor staging for lymphomas. NB: It is not guaranteed that data from the individual t_img, n_img and m_img variables have been combined into this variable, so they should be used in parallel.	TEXT	0, 1, 1A, 1A1, 1A2, 1B, 1B1, 1B2, 1C, 1E, 2, 2A, 2A1, 2A2, 2B, 2C, 2E, 2S, 3, 3A, 3B, 3C, 3E, 3S, 4, 4A, 4B, 4C, 4S, 6, ? U, X
System used to record imaging stage at diagnosis	stage_img_system	Version of the TNM classification of malignant cancers used to stage the tumour for the imaging TNM values	NUMBER	5 = 5th, 6 = 6th, 7 = 7th, 8 = 8th, 20 = UICC 5, 21 = UICC 6, 22 = UICC 7, 23 = AJCC 7, 24 = Unknown, 25 = UICC 8, 26 = AJCC 8, 27 = ENETS 2007
T stage (pathology)	t_path	The UICC code which classifies the size and extent of the primary tumour based on the evidence from a pathological examination	TEXT	0, 1, 1A, 1B, 1C, 1a, 1a1, 1a2, 1b, 1b1, 1b2, 1c, 1d, 1mi, 2, 2A, 2B, 2a, 2a1, 2a2, 2b, 2c, 2d, 2s, 3, 3A, 3a, 3b, 3c, 3d, 3s, 4, 4B,

				4a, 4b, 4c, 4d, 4e, A, IS, S, T1, T2, T3, TA, X, a, is
N stage (pathology)	n_path	The UICC code which classifies the absence or presence and extent of regional lymph node metastases based on the evidence from a pathological examination	TEXT	+, -, 0, 00, 1, 11, 1B, 1a, 1b, 1c, 1mi, 2, 2a, 2b, 2c, 3, 3a, 3b, 3c, 4, 9, N0, N1, N2, X
M stage (pathology)	m_path	The UICC code which classifies the absence or presence of distant metastases based on the evidence from a pathological examination	TEXT	0, 1, 1a, 1b, 1c, 1e, 2, 3, 4, 9, X, blank.
Pathological stage at diagnoses	stage_path	Pathologial stage at diagnosis. Combination of pathological T, N and M in “t_path”, “n_path” and “m_path”. Includes Ann Arbor staging for lymphomas. NB: It is not guaranteed that data from the individual t_path, n_path and m_path variables have been combined into this variable, so they should be used in parallel.	TEXT	0, 0A, 0IS, 1, 1A, 1A1, 1A2, 1B, 1B1, 1B2, 1C, 1E, 2, 2A, 2B, 2C, 2E, 3, 3A, 3B, 3C, 3E, 4, 4A, 4B, 4C, 5, 6, ?, U, X, blank
System used to record pathological stage at diagnosis	stage_path_system	Version of the TNM classification of malignant cancers used to stage the tumour for the pathological TNM values	NUMBER	5 = 5th 6 = 6th 7 = 7th 8 = 8th 20 = UICC 5 21 = UICC 6 22 = UICC 7 23 = AJCC 7 24 = Unknown 25 = UICC 8 26 = AJCC 8 27 = ENETS 2007

Stage (pathological pre-treatment)	stage_path_pretreated	Pathological stage at diagnosis recorded prior to treatment	TEXT	Y = Yes, X = No
Charlson co-morbidity score over 2 years	chrl_tot_27_03	Total Charlson co-morbidity score for a 2 year period. Lookback of 27 to 3 months prior to diagnosis	NUMBER	Diagnostic data for co-morbidities is derived from Admitted Care HES records matched at a patient level to cancer registration records, using the methodology published by Quan et al, Medical Care 43 1130-1139 (2005). Where recorded values range from 0 onwards
Charlson co-morbidity score over 6 years	chrl_tot_78_06	Total Charlson co-morbidity score for a 6 year period. Lookback of 78 to 6 months prior to diagnosis,	NUMBER	Diagnostic data for co-morbidities is derived from Admitted Care HES records matched at a patient level to cancer registration records, using the methodology published by Quan et al, Medical Care 43 1130-1139 (2005). Where recorded values

				range from 0 onwards
Oestrogen receptor status	er_status	Oestrogen receptor status of the tumour. These data are mainly recorded on invasive breast cancer C50x and DCIS (D05x).	TEXT	N = negative, P = positive, X = not performed
Oestrogen receptor score	er_score	Oestrogen receptor score of the tumour. These data are restricted to invasive breast cancer C50x and DCIS (D05x)	TEXT	ER Allred score (range 0, 2-8)
Progesterone receptor status	pr_status	Progesterone receptor status of the tumour. These data are mainly recorded on invasive breast cancer C50x and DCIS (D05x).	TEXT	N = negative, P = positive, X = not performed
Progesterone receptor score	pr_score	Progesterone receptor score of the tumour. These data are restricted to invasive breast cancer C50x and DCIS (D05x).	TEXT	ER Allred score (range 0, 2-8)

HER2 status	her2_status	Human Epidermal Growth Factor Receptor 2 status of the tumour. These data are mainly recorded on invasive breast cancer C50x and DCIS (D05x).	TEXT	N = negative, P = positive, X = not performed
NPI score	npi	Nottingham Prognostic Indicator score (not the derived stage) for prognosis following surgery for breast cancer. Calculated from tumour size, grade and lymph node involvement. Values >2 and <20 are accepted as valid. This data is restricted to invasive breast cancer C50x and DCIS (D05x)	NUMBER	Number (two decimal places) or blank

Dukes' stage	dukes	Used for colorectal cancer. Dukes stage of disease at diagnosis, based on pathological evidence but upgraded to Dukes D if clinical evidence of metastasis. This data is mainly recorded on invasive colorectal tumours (C18-21x).	TEXT	A = Dukes' A: Tumour confined to wall of bowel, nodes negative B = Dukes' B: Tumour penetrates through the muscularis propria to involve extramural tissues, nodes negative C1 = Dukes' C1: Metastases confined to regional lymph nodes (node/s positive but apical node negative) C2 = Dukes' C2: Metastases present in nodes at mesenteric artery ligature (apical node positive) D = Dukes D: Metastatic spread outside the operative field 99 = Not Known
FIGO stage	figo	Staging system for tumours of the female reproductive system. Stage 1 – Confined, Stage 2 – Local, Stage 3 – Distant, Stage 4 – Involving other organs. This data is mainly recorded on invasive gynaecological cancers (ICD 10 C51x to C57x) or retroperitoneum and peritoneum (C48x), as well as neoplasm of uncertain behaviour of ovary (D39x).	TEXT	0, 1, 1a, 1a1, 1a2, 1b, 1b1, 1b2, 1c, 1c1, 1c2, 1c3, 2, 2a, 2a1, 2a2, 2b, 2c, 3, 3a, 3b, 3c, 3c1, 3c2, 4, 4a, 4b, I, IA, IA1, IA2, IB, IB1, IB2, IC, II, IIA, IIA2, IIB, IIC, III, IIIA, IIIB, IIIC, IIIC1, IIIC2, IV, IVA, IVB, blank
Gleason primary pattern	gleason_primary	Used for prostate cancer (C61x) - the grade that comprises most of the tumour volume is called the "primary pattern"	NUMBER	1-5, 8 = not applicable
Gleason secondary pattern	gleason_secondary	Used for prostate cancer (C61x) - if additional grades present, the highest grade (biopsy) or the	NUMBER	1-5, 8 = not applicable

		second most extensive grade (TURP and radicals). If none present, primary and secondary grades are the same.		
Gleason tertiary pattern	gleason_tertiary	Used for prostate cancer (C61x) - value of any different third grade in addition to the primary and secondary grades	NUMBER	1-5, 8 = not applicable
Gleason combined	gleason_combined	Used for prostate cancer (C61x) - combined Gleason primary and secondary scores (+/- tertiary). The worst score is recorded, if both scores are the same but have different primary and secondary grades, the score with the highest primary grade is recorded first, i.e. 4+3=7 would be recorded over 3+4=7.	NUMBER	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, blank
Screen detected	screendetected	Whether or not the tumour was detected by a screening programme. Applicable to cancers identified through the NHS Cancer Screening Programmes only.	TEXT	N = No, Y = Yes, 8 = Not applicable, 9 = Not known
Screening status	screeningstatuscosd_code	Screening status. Applicable to cancers identified through the NHS Cancer Screening Programmes only.	TEXT	1 = screen-detected, 2 = interval cancer, 4 = lapsed attender, 5 = never attended, 6 = never invited, 7= Other, 9 = not known, NM = not mapped
Screening status description	screeningstatuscosd_name	Description of screening status	TEXT	
Screening status (detailed)	screeningstatusfull_code	The value of the sub-classification of the screening flag. Populated when the screening status is “Other”. For breast screening service: www.cancerscreening.nhs.uk/breastscreen/publications/nhsbsp62.pdf (page 4, section 2). For cervical screening service see: www.cancerscreening.nhs.uk/cervical/publications/nhscsp28.pdf (page 37). Applicable to cancers identified through the NHS Cancer Screening	TEXT	

		Programmes only.		
Description of full detailed screening status	screeningstatusfull_name	Description of the full detailed screening staturs.	TEXT	
Laterality	laterality	Identifies the side of the body for a tumour relating to paired organs within a patient, based on the evidence from a pathological examination. The data field is used to differentiate tumours in paired organs (and define the site of the pathology/cytology specimen). For paired sites, for e.g. the tonsils, if there is a tumour in one side, the laterality of that side, left or right, is recorded. For some paired sites, if there are tumours in both sides then two tumours are registered, one a left and the other a right. If there is a tumour in both sides (and they have other factors such as morphology the same) then only one registration is made, and the laterality is coded as bilateral. If the site of the primary cancer is not part of a pair, then laterality is coded as not applicable.	TEXT	L = Left, R = Right, M = Midline, B = Bilateral, 8 = Not applicable, 9 = Not Known
Diagnosis death certificate only	dco	Whether the tumour was registered from a death certificate only	TEXT	Y = Yes, N = No
Vital Status	vitalstatus	Records whether the patient is currently alive or deceased at the time of the snapshot.	TEXT	A =Alive, D =Dead, X =Exit posting.
Vital Status Date	vitalstatusdate	The last known date of vital status	DATE	ddmmyyyy
Date of first recorded hospitalisation	first_hosp_date	Date of first recorded hospitalisation. Available for 2008 onwards.	DATE	ddmmyyyy
Catchment area code	creg_code	Code for the cancer registry catchment area the patient was resident in when the tumour was diagnosed. Note that this variable cannot be provided alongside “region” variable of primary care	TEXT	Y0801=Thames Cancer Registry Y0201=Northern & Yorkshire Cancer Registry & Information Service Y0301=Trent Cancer Registry

		(i.e. if creg_code is provided, practice region will be redacted).		Y1201=West Midlands Cancer Intelligence Unit Y0401=Eastern Cancer Registration & Information Centre Y1701=North West Cancer Intelligence Service Y1001=South West Cancer Intelligence Service Y1101=Welsh Cancer Intelligence & Surveillance Unit Y0901=Oxford Cancer Intelligence Unit Z9999=blank
Catchment area name	creg_name	Description of the code provided in creg_code. Note that this variable cannot be provided alongside “region” variable of primary care (i.e. if creg_code is provided, practice region will be redacted).	TEXT	
Multifocal[1]	multifocal	Multifocal tumours are defined as discrete tumours apparently not in continuity with other primary cancers originating in the same site or tissue.	TEXT	N= No, Y = Yes, 8 = Not applicable, 9 = Not known
Clark's stage for skin cancer	clarks	Used for melanoma of the skin. Data is restricted to Malignant melanoma of skin (C43x) and Other malignant neoplasms of skin (C44x)	TEXT	1 = melanoma in situ: melanoma cells are only in the epidermis. 2 = melanoma cells in the papillary dermis. 3 = melanoma cells throughout papillary dermis and touching reticula. 4 = melanoma has spread into the reticular or deep dermis. 5 = melanoma has grown into the subcutaneous fat.
Breslow thickness of	breslow	Breslow thickness is used in staging melanoma of	TEXT	Number or range, x, or blank

[1] variables highlighted by Public Health England as those with potential data quality issues (multifocal, clarks, breslow, exicisionmargin)

tumour		the skin and is related to thickness of the tumour. Measured in millimetres to the nearest 0.01mm. Can be a number or range, x, or blank (e.g. ≤1mm, 1-2mm). Data is restricted to Malignant melanoma of skin (C43x) and Other malignant neoplasms of skin (C44x)		
Excision margin	excisionmargin	Whether the surgical excision margin finding was clear of the tumour and if so, by how much	TEXT	01 = Excision margins are clear (distance from margin not stated) 02 = Excision margins are clear (tumour >5mm from the margin) 03 = Excision margins are clear (tumour >1mm but less than or equal to 5mm from the margin 04 = Tumour is less than or equal to 1mm from excision margin, but does not reach margin 05 = Tumour reaches excision margin 06 = Uncertain 07 = Margin not involved =>1mm 08 = Margin not involved <1mm 09 = Margin not involved 1-5mm 98 = Not applicable 99 = Not Known

3. Treatment

Column description	*Column name*	*Details*	*Field Type*	*Valid Content*
CPRD patient Identifier	e_patid	Unique patient identifier based on CPRD primary care data – pseudonymised.	ID	Number
CR patient Identifier	e_cr_patid	Unique patient identifier based on NCRAS data - pseudonymised. In some cases the same person may have multiple patient IDs. Patient IDs will be retained even after two patient records are found to be the same person.	ID	Number
CR tumour identifier	e_cr_id	Unique tumour identifier based on NCRAS data – pseudonymised.	ID	Number
Number of tumours	number_of_tumours	Number of tumours affected by this event	NUMBER	
Type of event	eventcode	Type of event	TEXT	01a = Surgery - curative, 01b = Surgery - not curative, 01z = Surgery - type unknown, 02 = Cytotoxic Chemotherapy, 03 = Hormone Therapy, 04 = Chemoradiotherapy, 05 = RT–Teletherapy, 06 = RT–Brachytherapy, 07 = Specialist Palliative Care, 08 = Active Monitoring, 09 = Non-Specialist Palliative Care, 14 = Anti Cancer Drug Regimen (other), 15 = Immunotherapy, 19 = Radioisotope Therapy,

				22 = Radiosurgery, 97 = Other Treatment, 98 = All treatment declined, 99 = Treatment unknown, CTX = CT–Other, IM = Imaging, RTX = RT-Other/NK.
Description of event type	eventdesc	Description of event type	TEXT	
Treatment date	eventdate	Date the treatment took place	DATE	ddmmyyyy
Treatment within six months	within_six_months_flag	Whether treatment was within six months of date of diagnosis	NUMBER	0 = No, 1 = Yes
Treatment after six months	six_months_after_flag	Whether treatment was after six months from date of diagnosis	NUMBER	0 = No, 1 = Yes
OPCS code	opcs4_code	Operations, procedures and interventions coded. Code format X000.	TEXT	OPCS4 code
OPCS name	opcs4_name	Name of operation, procedure or intervention	TEXT	
Radiotherapy code	radiocode	Radiotherapy type	TEXT	1 = External Beam, 2 = Intracavity or Interstitial, 3 = 1 + 2, 4 = Radioactive Isotopes, 5 = 1 + 4, 8 = Other, B = Brachytherapy, X = Unknown / Inapplicable.
Radiotherapy code description	radiodesc	Description of radiotherapy type	TEXT	
Imaging code	imagingcode	Imaging code	TEXT	1 = Standard Radiography; 1A = Chest X-Ray; 1B = Sinus X-Rays; 1C = Mastoid Views; 1D = Orthopantomogram (OPG); 1E = Skull Base X-Rays; 1F = Angiography; 1G = Intravenous Urography; 1H =

				Retrograde Urography; 1J = Inferior Vena Cavography; 1K = Bone Angiography; 1L = Soft Tissue Angiography. 2 = CT Scan with unspecified contrast; 2A = CT scan with contrast; 2B = CT scan without contrast. 3 = MRI scan with unspecified contrast; 3A = MRI scan with contrast; 3B = MRI scan without contrast; 3C = MRI scan cholangiography. 4 = PET Scan. 5 = Ultrasound; 5A = Transabdominal ultrasound; 5B = Transvaginal ultrasound; 5C = Doppler ultrasound; 5D = Transrectal ultrasound; 5E = Endoscopic ultrasound; 5F = Laparoscopic ultrasound. 6 = Nuclear medicine imaging; 6A = Radio-isotope bone scan; 6B Other radio-isotope; 6C = Ventilation/Perfusion scan. 7 = Mammography. 8 = Barium; 8A = Barium Enema; 8B = Barium Swallow. 9 = Lymphoscintigraphy. 99 = Other. C02C = Virtual Colonoscopy. C08U = Urography (IV and retrograde). C09X = Intervention radiography. CXXX = Other (COSD).

Imaging code description	imagingdesc	Description of Imaging code	TEXT	
Imaging site	imagingsite	Site on body where imaging occurred	TEXT	Valid four digit OPCS4 code. Code Format: X000.
Lesion size	lesionsize	The size in mm of the diameter of a lesion, largest if more than one, if the histology of a site proves to be invasive.	NUMBER	Number or blank
Chemotherapy drug name(s)	chemo_all_drug	Chemotherapy drug name(s)	TEXT	
Name/acronym of known drug combinations	chemo_drug_group	Name/acronym of known drug combinations	TEXT	Please note this is a non-mandated text field that may not be complete or contain the specific drug name.