Public Health England

Protecting and improving the nation's health

National Cancer Intelligence Network

Linking English Cancer Patient Experience Surveys 2010, 2011/12 and 2013 with cancer registration data

A technical summary

Produced in partnership with Macmillan Cancer Support and Cancer Research UK

About Public Health England

Public Health England exists to protect and improve the nation's health and wellbeing, and reduce health inequalities. It does this through world-class science, knowledge and intelligence, advocacy, partnerships and the delivery of specialist public health services. PHE is an operationally autonomous executive agency of the Department of Health.

Public Health England
Wellington House
133-155 Waterloo Road
London SE1 8UG
Tel: 020 7654 8000
www.gov.uk/phe
Twitter: @PHE_uk
Facebook: www.facebook.com/PublicHealthEngland

Prepared by: National Cancer Intelligence Network
For queries relating to this document, please contact: NCINenquiries@phe.gov.uk

Published October 2015
PHE publications gateway number: 2015388

This document is available in other formats on request. Please call 020 7654 8158 or email enquiries@ncin.org.uk

The intelligence networks

Public Health England operates a number of intelligence networks, which work with partners to develop world-class population health intelligence to help improve local, national and international public health systems.

National Cancer Intelligence Network

The National Cancer Intelligence Network (NCIN) is a UK-wide initiative, working to drive improvements in standards of cancer care and clinical outcomes by improving and using the information collected about cancer patients for analysis, publication and research.

National Cardiovascular Intelligence Network

The National Cardiovascular Intelligence Network (NCVIN) analyses information and data and turns it into meaningful timely health intelligence for commissioners, policy makers, clinicians and health professionals to improve services and outcomes.

National Child and Maternal Health Intelligence Network

The National Child and Maternal Health Intelligence Network provides information and intelligence to improve decision-making for high-quality, cost-effective services. Its work supports policy makers, commissioners, managers, regulators, and other health stakeholders working on children's, young people's and maternal health.

National Mental Health, Dementia and Neurology Intelligence Network

The National Mental Health Intelligence Networks (NMHDNIN) brings together the distinct National Mental Health Intelligence Network, the Dementia Intelligence Network and the Neurology Intelligence Network under a single programme. The Networks work in partnership with key stakeholder organisations. The Networks seeks to put information and intelligence into the hands of decision makers to improve mental health and wellbeing, support the reduction of risk and improve the lives of people living with dementia and improve neurology services.

National End of Life Care Intelligence Network

The National End of Life Care Intelligence Network (NEoLCIN) aims to improve the collection and analysis of information related to the quality, volume and costs of care provided by the NHS, social services and the third sector to adults approaching the end of life. This intelligence will help drive improvements in the quality and productivity of services.

Contents

Executive summary

This technical report provides a description of the processes used and considerations made when linking the English Cancer Patient Experience Survey (CPES) data with the cancer registration data in PHE's Cancer Analysis System (CAS).

The record-level linkage will enrich the national cancer dataset available to analysts and researchers. It will enable a better understanding of patient care and the exploration of the relationships between patient experience and outcomes.

Introduction

The National Cancer Patient Experience Survey (CPES), commissioned by NHS England through Quality Health is a survey sent to cancer patients who have been discharged within a specified period from one of the acute and specialist NHS trusts in England that provide adult acute cancer services. The survey aims to collect information from patients about their cancer journey from their initial GP visit prior to diagnosis, through diagnosis and treatment to the ongoing management of their cancer.

This work links the CPES datasets to the core cancer registration data, allowing further analysis to increase our understanding about the patients' experience. This linkage follows the procedures set out in the data sharing agreement between PHE, NHS England and Quality Health.

This report covers the first stage of the project: to link each year of the CPES data to the tumour (AV_tumour) table in the Cancer Analysis System (CAS) in order to create merged datasets for analytical analyses. The next stage of the project is to undertake descriptive analysis of this linkage. Once this is published, the linked datasets will be available to researchers through the Office for Data Release.

For data access information, please see the section 'Where to find the linked data'.

PHE will continue to coordinate an analytical work programme between its partners and stakeholders for the linked CPES data.

Datasets

CPES

Currently three iterations of the CPES data are available in CAS – 2010, 2011/12 and 2013. Subsequent years will be added as both CPES and cancer registration data become available.

The questionnaire contains around 70 questions (74 in 2010, 77 in 2011/12 and 79 in 2013), excluding demographic questions. These cover the patient journey from initial GP visit prior to diagnosis, through diagnosis and treatment to the ongoing management of cancer.

The majority of the questions have stayed the same across the three surveys, however, there have been changes in question numbering and questions were added in the later years.

Each completed survey contains the recorded answers from patients and also information provided by the NHS trusts including the patient's NHS number, tumour site (ICD-10 code), gender, ethnicity and postcode of residence.

Variables such as the NHS number and tumour site enable record level linkage of the CPES datasets to the cancer registration data.

The full list of variables for each yearly CPES dataset have been listed in the appendix section and the methods for linking each survey are described in the methodology section.

Cancer registration data

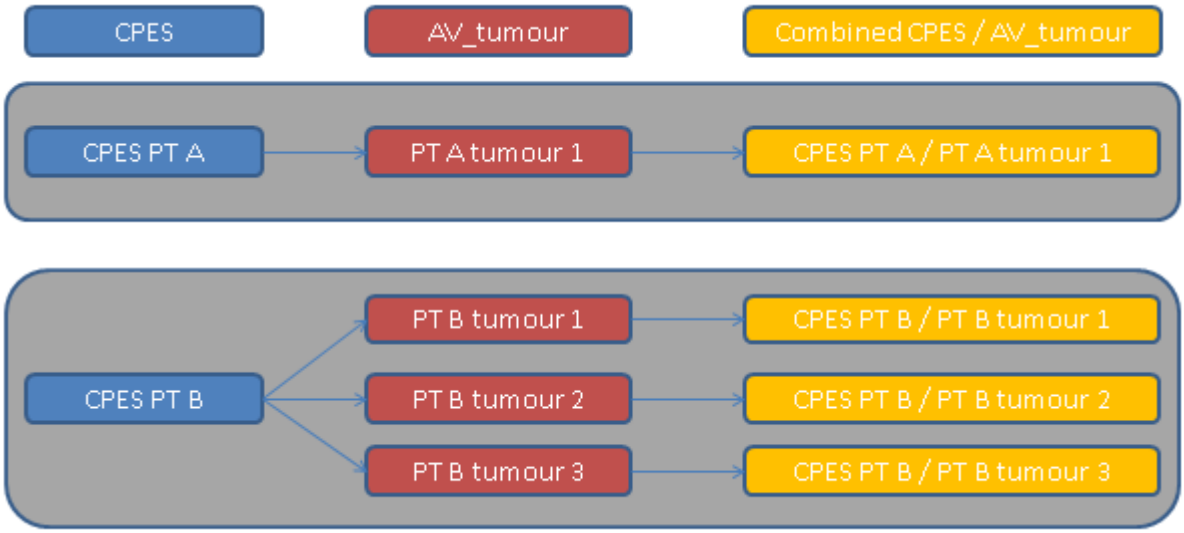
A snapshot of the cancer registration data in CAS is generated once a month. The table AV_tumour in the February 2015 snapshot (CAS1502) was used to link to the CPES datasets. This table contains record level data for each tumour diagnosed. Therefore, a patient diagnosed with more than one tumour will appear more than once in this table. The information regarding each tumour includes tumour site, date of diagnosis and other tumour related variables alongside patient characteristics such as gender, age and ethnicity.

The full list of variables available in the AV_tumour table can be found in the appendices.

CPES and CAS1502

The flow diagram below (figure 1) shows how records are initially linked at the patient-tumour level from CPES to the AV_tumour table in CAS 1502.

Figure 1. Examples of matching records between CPES and AV_tumour table



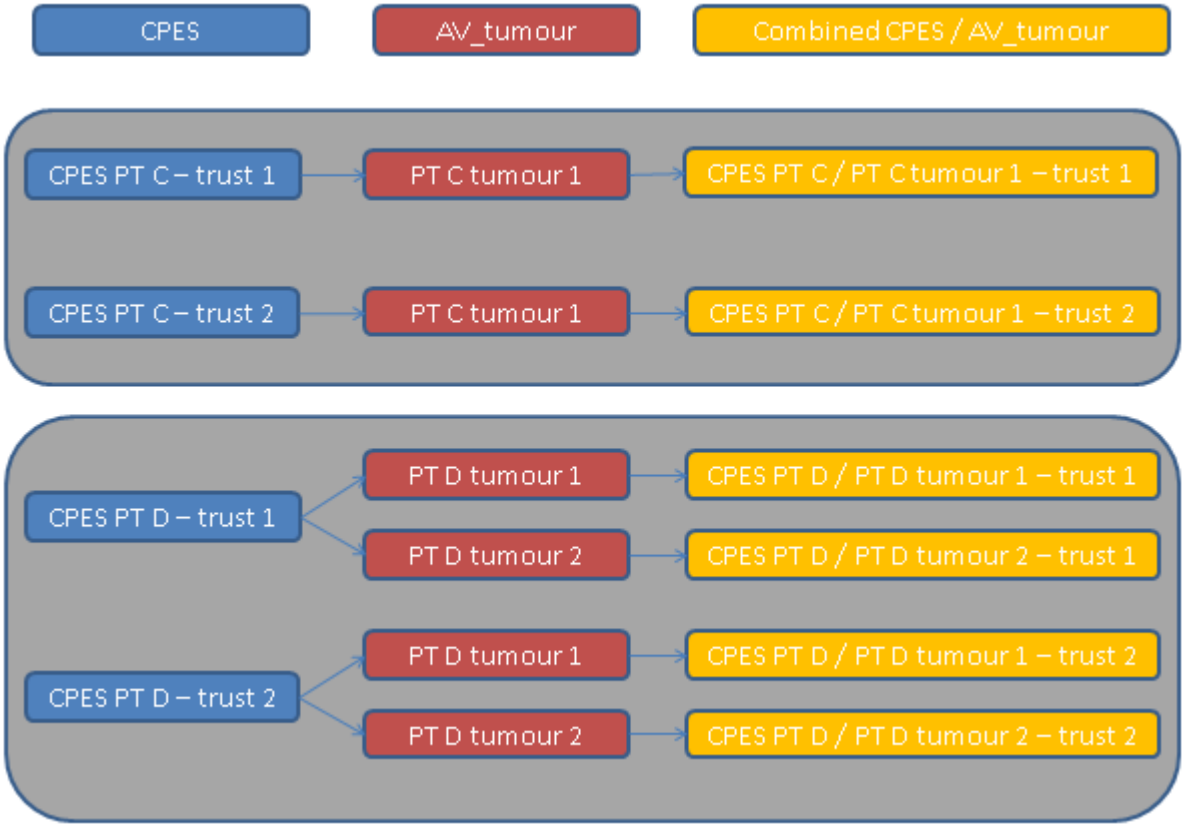
Here patient A has been linked by NHS number from the CPES dataset to the AV_tumour table where there is only one tumour record for the relating NHS number. This results in only one record in the merged dataset. However, patient B has three records in the AV_tumour table relating to their NHS number, this leads to the single CPES record being mapped onto each of the three tumour records. This generates multiple records for one NHS number in the merged dataset, with identical CPES entries but different AV_tumour table entries.

To address this issue a process of matching and data cleaning using the ICD-10 codes in both CPES and CAS has been applied, to ensure that the tumour is linked to the corresponding CPES record. This is described in detail in the methods section.

There were also instances of multiple survey responses from the same patients in the 2010 and 2011/12 CPES surveys. This was where patients were linked to more than one hospital trust for the same cancer and were approached separately to participate in the survey. This issue was resolved from the 2013 survey onwards.

Figure 2 below shows how a patient can be associated with more than one trust in CPES and how this gets linked with the AV_tumour table.

Figure 2. Examples of multiple trust records in CPES being linked to AV_tumour table



Here patient C has two CPES records from different NHS trusts but only one tumour record in the AV_tumour table. This means that when the records are linked by NHS number we are left with two records in the linked dataset. These records will have identical AV_tumour entries, but will have different CPES entries.

Patient D has also been reported by two different NHS trusts and they each also have two different tumour records in the AV_tumour table. When linked by NHS number, each CPES entry has been linked to each tumour record as described in figure 1. This leads to a total of four linked records for this NHS number.

The method used to address the multiple trust issue is described in the next section.

Methods

Dataset linkage via NHS number

In order to keep the integrity of the CPES linkage and for the purpose of analysis the linkage was done in a way to ensure no CPES records were deleted from the final dataset, even if there was no match in CAS1502.

The record linkage was based on individual identifiers that matched across both datasets. The patients' NHS number was the identifier used to link records in both CPES and cancer registration datasets. Records with an identical NHS number were considered matched at patient level. A variable was created to reflect the outcome of the matching via NHS number called '_MERGE'. It is coded '0' for records with no match between the datasets, and '1' for records with a match in NHS number between the datasets. A list of the new flag variables created can be found in the appendices along with their corresponding value formats.

The datasets resulting from the linkage of 2010, 2011/12 and 2013 CPES and CAS1502 have the following characteristics:

Table 1. Records linked and not linked via NHS number

	2010	2011/12	2013
(a) CPES records – N	67,713	71,793	68,737
(b) CPES records linked – n (% of a)	66,179 (98%)	70,756 (99%)	68,076 (99%)
CPES Single Trust records linked – n (% of b)	65,787 (99%)	70,270 (99%)	68,076 (100%)
CPES Double Trust Patients linked – n (% of b)	392 (1%)	486 (1%)	0 (0%)
(c) CPES records not linked – n	1,142 (2%)	551 (1%)	661 (1%)
CPES records with an NHS number that did not linked with CAS1502 – n (% of a)	483 (1%)	542 (1%)	630 (1%)
CPES records with no match due to no/not valid NHS number– n (% of a)	659 (1%)	9 (0%)	31 (0%)

Tumour linkage

The following criteria were then used to match records at patient-tumour level, relevant flags were created for each record based on the precision of the match on ICD-10 codes.

1. A match was established if the ICD-10 4-digit code in CPES and CAS1502 (AV_tumour) were identical. This match was flagged via the variable 'FLAGICD-104'
2. For the remaining non-matched records, a match was established between ICD-10 code with 3 digits in CPES and CAS1502 (AV_tumour). This match is flagged in the variable 'FLAGICD-103'

A summary variable called 'TUMOUR_LINKFLAG_COMBINED' was created to indicate whether the match was at ICD-10 4-digit or 3-digit level.

All records were then 'cleaned' to select the registration nearest to the time of the survey date. After evaluating the distribution of time between diagnosis and discharge, there were a number of patients that appeared to have been diagnosed after their discharge date, with a large proportion of these patients being in the time window of 1-30 days after their discharge. For this reason, patients who were diagnosed over 30 days after their discharge date in the CPES dataset were not indicated as a match. This was applied to help remove any records that were deemed 'duplicates' (by duplicate we mean records that were identical for a patient on given variables of interest, and in this case a patient that had two or more records that were a match in ICD-10 4-digit or 3-digit level), but also to remove any record that matched, but had a date of diagnosis more than 30 days after discharge to minimise the possibility of error.

Records that were not matched were retained and flagged accordingly in the linked dataset.

The results of the linkage of 2010, 2011/12 and 2013 CPES and CAS1502 based on both the patient's NHS number and ICD-10 code (4 or 3 digits) are shown in Table 2.

Table 2. Patients linked via NHS number and ICD-10 code 4 digits and 3 digits

	2010	2011/12	2013
(a) CPES patients – N	**67,321**	**71,306**	**68,737**
(b) CPES records – N	**67,713**	**71,793**	**68,737**
(c) Records in CAS1502 linked via NHS number – n	89,586	94,148	88,828
Records matching on ICD-10 4 digits – n (% of c)	19,044 (21%)	31,672 (34%)	30,369 (34%)
Records matching on ICD-10 3 digits – n (% of c)	35,663 (40%)	26,434 (28%)	25,745 (29%)
Records matching on ICD-10 4 and 3 digits including duplicates – n (% of c)	54,707 (61%)	58,106 (62%)	56,114 (63%)
Patients matching on ICD-10 4 and 3 digits excluding duplicates – n (% of a)	**51,792 (77%)**	**55,686 (78%)**	**54,355 (79%)**

Other variables were also created in order to help with record linkage. For a list of the variables created to support this project please see the appendices.

Tumour linkage of similar records

In order to increase the matching yield, an additional variable was created based on related tumour types in the two datasets, following consultation with clinical and research experts in the field. Records of patients recorded as having related tumour types, which included examples such as breast cancer, coded as 'C50' in one dataset and 'D05' (in-situ) in the other, were flagged.

For these flagged tumours a more stringent rule in terms of time between diagnosis and discharge was necessary given the large variance across this group. These cases were considered a match if the time between diagnosis and discharge was within the mean time frame for the most relevant tumour. So for instance, the average length of time between diagnosis and discharge among matched breast cancer patients was 546 days in 2010 (considering just the positive time interval values). Therefore, cases that were coded 'C50' in CPES and 'D05' in the AV_tumour table, and vice-versa, with a time difference from diagnosis and discharge date between -30 days and 546 days were considered a related match and flagged. This flag variable is called 'FLAG_RELATED_MATCH' (see list of variables in appendices).

There was an increase in the number of matched observations by including related tumours as summarised in Table 3:

Table 3. Patients linked via NHS number, ICD-10 code 4 digits and 3 digits match, and further including some non-identical ICD-10 codes (with high occurrence)

	2010	2011/12	2013
(a) CPES patients – N	**67,321**	**71,306**	**68,737**
Patients matching on ICD-10 4 and 3 digits excluding duplicates – n (% of a)	51,792 (77%)	55,686 (78%)	54,355 (79%)
Additional patients matching of related tumours– n (% of a)	4,149 (6%)	4,233 (6%)	3,551 (5%)
Patients matching on ICD-10 4 and 3 digits and additional related tumours[1] – n (% of a)	55,941 (83%)	59,919 (84%)	57,906 (84%)

[1] related tumour refers to tumours that were flagged in variable 'FLAG_RELATED_MATCH' as per the description above.

Merging CPES 2010, 2011/12 and 2013

When merging the three data files together, special care needs to be taken due to the fact that different CPES surveys have assigned different question numbers to similar questions. A rename procedure needs to take place prior to the merging of the datasets. For a guideline of the dataset fields across surveys and a suggestion of new field names, please see the link in the section below or request the document through enquiries@ncin.org.uk.

Where to find the linked data

Information on the linked database and a data dictionary for the CPES variables are available on the CPES webpage of the NCIN website:

www.ncin.org.uk/cancer_information_tools/cancer_patient_experience

For all enquiries regarding the linked CPES datasets and analysis programmes please contact NCIN using the following email address:

NCINenquiries@phe.gov.uk

For all data access requests please contact the Office for Data Release (ODR) using the following email address:

ODR@phe.gov.uk

Project team and acknowledgements

This technical report and the data linkage were prepared by Isabella Carneiro and James Charnock who are members of the CRUK-NCIN partnership and the Macmillan-NCIN partnership, respectively.

The National Cancer Intelligence Network wishes to acknowledge and thank all of those who have contributed to the content and production of this publication, and without whom the work would not have been possible.

Special thanks to Dr Mick Peake, and Dr Yoryos Lyratzopoulos who acted as the clinical and research advisor, respectively. We are also thankful to the CPES Consortium that provided support, guidance and resources as this project developed.

We would like to acknowledge the essential work of the National Cancer Registration Service and Quality Health, without whom we would not have any cancer registration or CPES data.

Appendix

Dataset variables, source and description

CPES 2010, 2011/12 and 2013 linkage to CAS1502

2010 (N=90,728)	2011/12 (N=94,699)	2013 (N= 89,549)
CPES variables		
CPES_TRUST_CODE	CPES_TRUST_CODE	CPES_TRUST_CODE
CPES_SERVICE_USER_RECORD_NUMBER	CPES_SERVICE_USER_RECORD_NUMBER	CPES_SERVICE_USER_RECORD_NUMBER
CPES_NHSNUMBER	CPES_NHSNUMBER	CPES_NHSNUMBER
CPES_DATE_OF_BIRTH	CPES_DATE_OF_BIRTH	CPES_DATE_OF_BIRTH
CPES_POSTCODE	CPES_POSTCODE	CPES_POSTCODE
CPES_GENDER	CPES_YEAR_OF_BIRTH	CPES_GENDER
CPES_YEAR_OF_BIRTH	CPES_SAMPLE_YOB_AGE_BAND	CPES_IMD_2007
CPES_ETHNICITY	CPES_DAY_OF_ADMISSION	CPES_IMD_QUINTILE
CPES_DAY_OF_ADMISSION	CPES_MONTH_OF_ADMISSION	CPES_ETHNICITY
CPES_MONTH_OF_ADMISSION	CPES_YEAR_OF_ADMISSION	CPES_DAY_OF_ADMISSION
CPES_YEAR_OF_ADMISSION	CPES_DAY_OF_DISCHARGE	CPES_MONTH_OF_ADMISSION
CPES_DAY_OF_DISCHARGE	CPES_MONTH_OF_DISCHARGE	CPES_YEAR_OF_ADMISSION
CPES_MONTH_OF_DISCHARGE	CPES_YEAR_OF_DISCHARGE	CPES_DAY_OF_DISCHARGE
CPES_YEAR_OF_DISCHARGE	CPES_LENGTH_OF_STAY	CPES_MONTH_OF_DISCHARGE
CPES_LENGTH_OF_STAY	CPES_ICD10_CODE	CPES_YEAR_OF_DISCHARGE
CPES_MAIN_SPECIALTY	CPES_ICD10_GROUP	CPES_LENGTH_OF_STAY
CPES_REFERRING_PCT	CPES_ETHNICITY	CPES_ICD10_CODE
CPES_ICD10_CODE	CPES_GENDER	CPES_ICD10_GROUP
CPES_MAIN_ICD10_CODE	CPES_MAIN_SPECIALTY	CPES_MAIN_SPECIALTY

CPES_ICD10_GROUP	CPES_REFERRING_PCT	CPES_REFERRING_PCT
CPES_PATIENT_CLASSIFICATION	CPES_DEPRIVATION_VALUE	CPES_PATIENT_CLASSIFICATION
CPES_HOSPITAL_SITE	CPES_PATIENT_CLASSIFICATION	CPES_YEAR_OF_BIRTH
CPES_DEPRIVATION_VALUE	CPES_HOSPITAL_SITE	CPES_SAMPLE_YOB_AGE_BAND
CPES_DATE_LOGGED	CPES_Q01	CPES_Q01
CPES_RESPONSE	CPES_Q02	CPES_Q02
CPES_RESPONSE_TYPE	CPES_Q03	CPES_Q03
CPES_Q01	CPES_Q04	CPES_Q04
CPES_Q02	CPES_Q05	CPES_Q05
CPES_Q03	CPES_Q06	CPES_Q06
CPES_Q04	CPES_Q07	CPES_Q07
CPES_Q05	CPES_Q08	CPES_Q08
CPES_Q06	CPES_Q09	CPES_Q09
CPES_Q07	CPES_Q10	CPES_Q10
CPES_Q08	CPES_Q11	CPES_Q11
CPES_Q09	CPES_Q12	CPES_Q12
CPES_Q10	CPES_Q13	CPES_Q13
CPES_Q11	CPES_Q14	CPES_Q14
CPES_Q12	CPES_Q15	CPES_Q15
CPES_Q13	CPES_Q16	CPES_Q16
CPES_Q14	CPES_Q17	CPES_Q17
CPES_Q15	CPES_Q18	CPES_Q18
CPES_Q16	CPES_Q19	CPES_Q19
CPES_Q17	CPES_Q20	CPES_Q20
CPES_Q18	CPES_Q21	CPES_Q21
CPES_Q19	CPES_Q22	CPES_Q22
CPES_Q20	CPES_Q23	CPES_Q23
CPES_Q21	CPES_Q24	CPES_Q24
CPES_Q22	CPES_Q25	CPES_Q25

CPES_Q23	CPES_Q26	CPES_Q26
CPES_Q24	CPES_Q27	CPES_Q27
CPES_Q25	CPES_Q28	CPES_Q28
CPES_Q26	CPES_Q29	CPES_Q29
CPES_Q27	CPES_Q30	CPES_Q30
CPES_Q28	CPES_Q31	CPES_Q31
CPES_Q29	CPES_Q32	CPES_Q32
CPES_Q30	CPES_Q33	CPES_Q33
CPES_Q31	CPES_Q34	CPES_Q34
CPES_Q32	CPES_Q35	CPES_Q35
CPES_Q33	CPES_Q36	CPES_Q36
CPES_Q34	CPES_Q37	CPES_Q37
CPES_Q35	CPES_Q38	CPES_Q38
CPES_Q36	CPES_Q39	CPES_Q39
CPES_Q37	CPES_Q40	CPES_Q40
CPES_Q38	CPES_Q41	CPES_Q41
CPES_Q39	CPES_Q42	CPES_Q42
CPES_Q40	CPES_Q43	CPES_Q43
CPES_Q41	CPES_Q44	CPES_Q44
CPES_Q42	CPES_Q45	CPES_Q45
CPES_Q43	CPES_Q46	CPES_Q46
CPES_Q44	CPES_Q47	CPES_Q47
CPES_Q45	CPES_Q48	CPES_Q48
CPES_Q46	CPES_Q49	CPES_Q49
CPES_Q47	CPES_Q50	CPES_Q50
CPES_Q48	CPES_Q51	CPES_Q51
CPES_Q49	CPES_Q52	CPES_Q52
CPES_Q50	CPES_Q53	CPES_Q53
CPES_Q51	CPES_Q54	CPES_Q54
CPES_Q52	CPES_Q55	CPES_Q55

CPES_Q53	CPES_Q56	CPES_Q56
CPES_Q54	CPES_Q57	CPES_Q57
CPES_Q55	CPES_Q58	CPES_Q58
CPES_Q56	CPES_Q59	CPES_Q59
CPES_Q57	CPES_Q60	CPES_Q60
CPES_Q58	CPES_Q61	CPES_Q61
CPES_Q59	CPES_Q62	CPES_Q62
CPES_Q60	CPES_Q63	CPES_Q63
CPES_Q61	CPES_Q64	CPES_Q64
CPES_Q62	CPES_Q65	CPES_Q65
CPES_Q63	CPES_Q66	CPES_Q66A
CPES_Q64	CPES_Q67	CPES_Q66B
CPES_Q65	CPES_Q68	CPES_Q66C
CPES_Q66	CPES_Q69	CPES_Q66D
CPES_Q67	CPES_Q70	CPES_Q66E
CPES_Q68	CPES_Q71	CPES_Q67
CPES_Q69	CPES_Q71_AGE_BAND	CPES_Q68
CPES_Q70A	CPES_Q72	CPES_Q69
CPES_Q70B	CPES_Q73	CPES_Q70
CPES_Q70C	CPES_Q74A	CPES_Q71
CPES_Q70D	CPES_Q74B	CPES_Q71_AGE
CPES_Q70E	CPES_Q74C	CPES_Q71_AGE_BAND
CPES_Q70F	CPES_Q74D	CPES_Q72
CPES_Q70G	CPES_Q74E	CPES_Q73
CPES_Q71	CPES_Q74F	CPES_Q74
CPES_Q72	CPES_Q74G	CPES_Q75A
CPES_Q73	CPES_Q75	CPES_Q75B
CPES_Q74	CPES_Q76	CPES_Q75C
CPES_Q74_OTHER	CPES_Q77	CPES_Q75D
		CPES_Q75E

		CPES_Q75F
		CPES_Q75G
		CPES_Q76
		CPES_Q77
		CPES_Q78
		CPES_Q79

CAS1502 variables
AVTUM_TUMOURID
AVTUM_PATIENTID
AVTUM_DIAGNOSISDATEBEST
AVTUM_DIAGNOSISYEAR
AVTUM_SITE_CODED
AVTUM_SITE_CODED_DESC
AVTUM_SITE_CODED_3CHAR
AVTUM_MORPH_CODED
AVTUM_BEHAVIOUR_CODED
AVTUM_BEHAVIOUR_CODED_DESC
AVTUM_CODING_SYSTEM
AVTUM_CODING_SYSTEM_DESC
AVTUM_HISTOLOGY_CODED
AVTUM_HISTOLOGY_CODED_DESC
AVTUM_SITE_ICD10_O2
AVTUM_SITE_ICD10_O2_3CHAR
AVTUM_MORPH_ICD10_O2
AVTUM_BEHAVIOUR_ICD10_O2
AVTUM_DCO
AVTUM_T_BEST
AVTUM_N_BEST
AVTUM_M_BEST

AVTUM_STAGE_BEST
AVTUM_STAGE_BEST_SYSTEM
AVTUM_GRADE
AVTUM_TUMOURSIZE
AVTUM_NODESEXCISED
AVTUM_NODESINVOLVED
AVTUM_LATERALITY
AVTUM_MULTIFOCAL
AVTUM_BASISOFDIAGNOSIS
AVTUM_BIRTHDATEBEST
AVTUM_AGE
AVTUM_FIVEYEARAGEBAND
AVTUM_SEX
AVTUM_POSTCODE_1
AVTUM_LSOA11_CODE
AVTUM_LSOA01_CODE
AVTUM_MSOA11_CODE
AVTUM_MSOA01_CODE
AVTUM_CCG_CODE
AVTUM_CCG_NAME
AVTUM_PCT_CODE
AVTUM_PCT_NAME
AVTUM_LAUA_CODE
AVTUM_LAUA_NAME
AVTUM_UTLA_CODE
AVTUM_UTLA_NAME
AVTUM_SCN_CODE
AVTUM_SCN_NAME
AVTUM_CNET_CODE
AVTUM_CNET_NAME

AVTUM_COUNTY_CODE
AVTUM_COUNTY_NAME
AVTUM_GOR_CODE
AVTUM_GOR_NAME
AVTUM_CREG_CODE
AVTUM_CREG_NAME
AVTUM_DIAGNOSISPROVIDER_CODE
AVTUM_DIAGNOSISPROVIDER_NAME
AVTUM_DIAGNOSISTRUST_CODE
AVTUM_DIAGNOSISTRUST_NAME
AVTUM_ONSID
AVTUM_NHSNUMBER_1
AVTUM_STATUSOFREGISTRATION
AVTUM_CENTRE
AVTUM_CENTRENAME
AVTUM_EXCISIONMARGIN
AVTUM_SCREENDETECTED
AVTUM_SCREENINGSTATUSCOSD_CODE
AVTUM_SCREENINGSTATUSCOSD_NAME
AVTUM_SCREENINGSTATUSFULL_CODE
AVTUM_SCREENINGSTATUSFULL_NAME
AVTUM_ER_STATUS
AVTUM_ER_SCORE
AVTUM_PR_STATUS
AVTUM_PR_SCORE
AVTUM_HER2_STATUS
AVTUM_NPI
AVTUM_DUKES
AVTUM_FIGO
AVTUM_CLARKS

AVTUM_BRESLOW
AVTUM_GLEASON_PRIMARY
AVTUM_GLEASON_SECONDARY
AVTUM_GLEASON_TERTIARY
AVTUM_GLEASON_COMBINED
AVTUM_T_IMG
AVTUM_N_IMG
AVTUM_M_IMG
AVTUM_STAGE_IMG
AVTUM_STAGE_IMG_SYSTEM
AVTUM_T_PATH
AVTUM_N_PATH
AVTUM_M_PATH
AVTUM_STAGE_PATH
AVTUM_STAGE_PATH_SYSTEM
AVTUM_STAGE_PATH_PRETREATED
AVTUM_DIAGNOSISDATE1
AVTUM_DIAGNOSISDATE2
AVTUM_DIAGNOSISDATEFLAG
AVTUM_FINANCIALYEAR
AVTUM_ETHNICITY_1
AVTUM_VITALSTATUS
AVTUM_DEATHDATEBEST
AVTUM_DEATHDATEFLAG
AVTUM_EMBARKATION
AVTUM_EMBARKATIONDATE
AVTUM_DATE_FIRST_EVENT
AVTUM_TRUSTCODE_FIRST_EVENT
AVTUM_TRUSTNAME_FIRST_EVENT
AVTUM_DATE_FIRST_SURGERY

AVTUM_TRUSTCODE_FIRST_SURGERY
AVTUM_TRUSTNAME_FIRST_SURGERY
AVTUM_CASCADE_INCI_FLAG

New variables created (labels) and values

Variable name	Values
_MERGE (Flag for observations matched on NHSnr between CPES and CAS1502)	0 'Records not matched' 1 'Records matched'
FLAGICD104 (FLAG for observations match on ICD10 4 digits between CPES and CAS1502)	0 'Records not matched' 1 'Records matched'
FLAGICD-103 (FLAG for observations match on ICD10 3 digits between CPES and CAS1502)	0 'Records not matched' 1 'Records matched'
TUMOUR_LINK_FLAG_COMBINED (Field showing observations matched on ICD10 4 and 3 digits, or not, between CPES and CAS1502)	0 'Records not matched' 1 'Records matched'
DIFFDIAGDISCH (Difference from data of diagnosis (CPES) and date of Discharge (CAS1502) in days)	Days
PATMATCH (Flag marking patients that were matched in NHSnr and most recent tumour)	0 ' Patient not matched' 1 ' Patient matched' 9999 ' Patient not matched in NHSnumber'
FINAL_UNIQUE (Field depicting unique patients, double records and no matches in CAS1502)	0 'Duplicate Record plus non-tumour matched records' 1 'Final Unique Patient' 99 'No match in NHSnumber'
FLAG_RELATED_MATCH (Flag for matching the not-equal ICD10 codes that were similar and could be matched)	1 'Records matched'
FLAG_RELATED_MATCH_UNIQ (Flag for unique observations matching the not-equal ICD10 codes that were similar and could be matched)	0 'Records not matched' 1 'Records matched'
FINAL_UNIQUE_EXTRA (Flag similar to Final Unique combined with higher yield by adding non-equal ICD10)	0 'Duplicate Record plus non-tumour matched records' 1 'Final Unique Patient' 99 'No match in NHSnumber'

Related ICD-10 Codes used for FLAG_Related_Match variable		
ICD-10 Code		**Base ICD-10 Code**
C67	D41	C67
C67	D09	C67
C83	C85	C83
C85	C82	C82
C50	D05	C50
C20	C19	C18, C19 and C20 (Mean)
C18	C19	C18, C19 and C20 (Mean)
C18	C20	C18, C19 and C20 (Mean)
C82	C83	C18, C19 and C20 (Mean)

C18 - Malignant neoplasm of colon **C19** - Malignant neoplasm of rectosigmoid junction **C20** - Malignant neoplasm of rectum **C50** - Malignant neoplasm of breast **C67** - Malignant neoplasm of bladder **C82** - Follicular [nodular] non-Hodgkin's lymphoma **C83** - Diffuse non-Hodgkin's lymphoma **C85** - Other and unspecified types of non-Hodgkin's lymphoma **D41** - Neoplasm of uncertain or unknown behaviour of urinary organs **D09** - Carcinoma in situ of other and unspecified sites **D05** - Carcinoma in situ of breast