



Public Health
England



CANCER
RESEARCH
UK

Protecting and improving the nation's health

Cancer analysis system: standard operating procedure

Defining the Secondary Care Diagnostic
Interval using AV_Tumour linked data

About Public Health England

Public Health England exists to protect and improve the nation's health and wellbeing, and reduce health inequalities. We do this through world-leading science, knowledge and intelligence, advocacy, partnerships and the delivery of specialist public health services. We are an executive agency of the Department of Health and Social Care, and a distinct delivery organisation with operational autonomy. We provide government, local government, the NHS, Parliament, industry and the public with evidence-based professional, scientific and delivery expertise and support.

This report has been produced as part of the Cancer Research UK/Public Health England partnership.

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

For queries relating to this document, please contact: NCRASenquiries@phe.gov.uk

© Crown copyright 2019

You may re-use this information (excluding logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0. To view this licence, visit [OGL](https://www.ogcl.gov.uk). Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

Published: May 2019
PHE publications
gateway number: GW-406

PHE supports the UN
Sustainable Development Goals



Contents

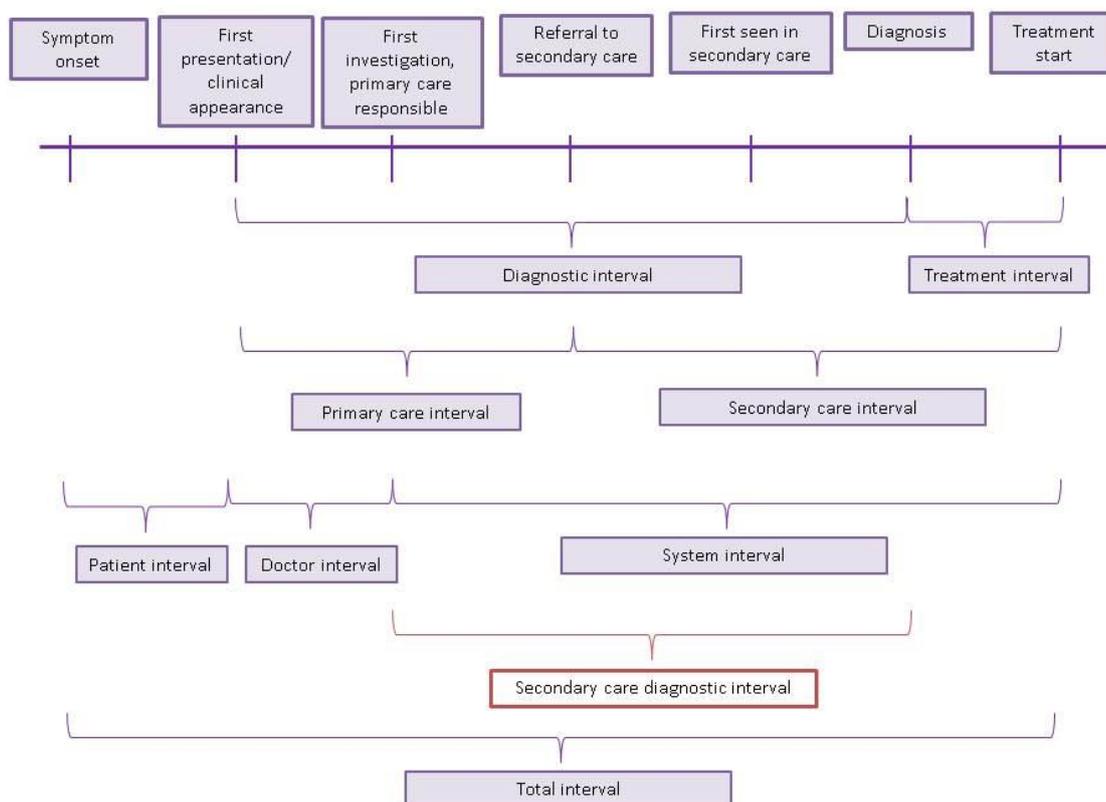
About Public Health England	2
Contents	3
Introduction	4
Summary of terms	6
Method	7
Appendix 1: cancer sites	11
Appendix 2: diagnostic test codes by cancer site	12
Appendix 3: example code	26

Introduction

The purpose of this Standard Operating Procedure (SOP) is to describe the method used to define the Secondary Care Diagnostic Interval (SCDI) using cancer registry data and linked datasets in the Cancer Analysis System (CAS). This allows for the generation of a measure of SCDI length (in days) for as many patients as possible across the majority of cancer sites. Cancer sites were selected for inclusion based on interest from stakeholders in combination with sites included in previously published [Routes to Diagnosis](#) and [Treatments](#) work. This method was used for recent CRUK-PHE partnership publications of SCDI work including the data visualisation tool 'Secondary Care Diagnostic Interval, 2014-2015' and forthcoming journal article 'Establishing population-based surveillance of diagnostic timeliness using linked cancer registry and administrative data for patients with colorectal and lung cancer' (both available here www.ncin.org.uk/about_ncin/scdi). This is the first National Cancer Registration and Analysis Service (NCRAS) methodology to calculate a diagnostic interval for all cancer patients in England.

The SCDI is defined as the period of time commencing when a patient is referred to or enters secondary care for investigations for a suspected cancer to when they receive a cancer diagnosis. The interval differs in its parameters (start and/or end point) from similar previously defined intervals (see figure 1).

Figure 1 – Secondary Care Diagnostic Interval in relation to other time intervals



Adapted from: Weller D, Vedsted P, Rubin G, Walter FM, Emery J, Scott S, et al. The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. *Br J Cancer* [Internet]. 2012;106. Available from: <http://dx.doi.org/10.1038/bjc.2012.68>

Secondary care is used as a start point because cancer registry data (and linked datasets) does not currently include comprehensive coverage of national primary care. Prior to the introduction of this methodology, only patients referred on the urgent referral for suspected cancer and patients diagnosed via screening had a recognised start point recorded through the Cancer Waiting Times (CWT) dataset, accounting for ~43% of all cancers diagnosed in 2015. For all other patients this methodology established a starting point using diagnostic testing, referral dates and secondary care appointment information.

This SOP is to be used where the analyst wishes to:

- identify a start date for the SCDI
- calculate the length of SCDI in days

This SOP exists to set a standard that can be followed to produce uniform and replicable results. Certain specific uses may require a different approach and should be discussed with the lead of the early pathway and diagnosis functional team.

The scripts (SQL, R) which accompany this SOP are in appendix 3. The code produces patient-level data with a SCDI start date and a SCDI length in days.

Summary of terms

To define the SCDI, it is necessary to define both a start and an end date. The start point is a composite of event types which are:

- diagnostic test in secondary care
- referral to secondary care appointment
- secondary care appointment

The full list of 25 cancer sites for which the SCDI can be calculated can be found in appendix 1. For each cancer site, where relevant, a list of relevant diagnostic tests was defined in consultation with clinicians (see appendix 2 for the full list by site). In the case of this SOP the term 'diagnostic test' refers to:

- diagnostic procedures: recorded in Admitted Patient Care (APC) Hospital Episode Statistics (HES) defined using OPCS-4 codes (more information [here](#))
- diagnostic radiological imaging: recorded in the Diagnostic Imaging Dataset (DID) – defined using SNOMED CT or NICIP codes (more information [here](#))

The end point was either the registry defined diagnosis date OR an MDT OR treatment referral/ start date if this occurred up to 28 days before the registry diagnosis date. The interval length is calculated by finding the difference in days between the start and end dates.

The SCDI is limited to a specific timeframe – the 6 months directly preceding diagnosis date (from 0 to 182 days inclusive) – to strike a balance between identifying the earliest occurring relevant secondary care diagnostic event while minimising the inclusion of events unrelated to the tumour diagnosis.

Method

Cohort and overall approach

Cancer registry data from AV_Tumour is used as the base to identify the cohort of patients. The datasets used to collate events relating to the starting point of the pathway are:

- AV_pathway which is a derived dataset comprised of events relating to various stages of the patient pathway - events of interest in this dataset were sourced from:
 - Cancer Waiting Times (CWT) which is a dataset used to support management and monitoring of cancer services in England
 - Routes to Diagnosis (RtD) derived dataset which is a derived dataset that assigns to national cancer registry patients one of 8 routes to diagnosis
- Diagnostic Imaging Dataset (DID) which is a dataset providing detailed information on diagnostic imaging, extracted from local Radiology Information Systems
- Hospital Episode Statistics (HES) which is a dataset containing details of admitted patient, outpatient, accident and emergency and adult critical care appointments in England

The scope of this SOP is tumours diagnosed from 2014-15 as DID data recorded before this time is known to be of a lower quality. This method can also be applied to tumour registrations in subsequent years, although the codes relating to diagnostic testing must be re-checked with clinical advisors as these may change over time.

Exclusions

Patients with a previous primary cancer of the same site (recorded in the period 2012-15) are excluded in order to avoid including data from a previous diagnosis. Additionally patients whose diagnosis was recorded via death certificate only, those with a missing NHS number and those with an unknown route to diagnosis are excluded, as calculating a SCDI for these patients is not possible. Males with gynaecological or breast cancer and females with prostate cancer are also excluded.

End point: diagnosis date – amended

The end point of the SCDI is a cancer diagnosis. All patients have a diagnosis date recorded in the national cancer registry (AV_Tumour) which is used in the first instance. However, the European Network of Cancer Registries (ENCR) hierarchy states that date of diagnosis is recorded as the date of most recent pathological confirmation (where available). Though this allows for international comparison it is less useful for the purposes of accurately calculating the SCDI as, in practice, patients may receive a clinical diagnosis

some time before pathological verification. In order to more accurately reflect clinical practice, if there is a record of either:

- a Multi-Disciplinary Team (MDT) meeting
- treatment referral
- treatment initiation

in the 28 days (0-28 days inclusive) preceding the registry diagnosis date, this date is used as the diagnosis date/ end point instead of the registry date (as these event types indicate the presence of a clinical diagnosis). In cases where one of the above listed event types occurred more than 28 days before the registry date there is less certainty that the data related to the same tumour, so the registry date is retained as the end point.

Start point: first event in secondary care

The start point reflects when a patient is referred into or enters secondary care. This method combines AV_Tumour cohort data with various linked datasets to establish one of any first events occurring in the 6 months (0-182 days inclusive) preceding a cancer diagnosis for each patient:

- urgent referral for suspected cancer
 - source: CWT
 - linked using: AV_Pathway derived dataset
 - AV_Pathway event_type*= 1
- secondary care appointment for suspected cancer
 - source: CWT
 - linked using: AV_Pathway derived dataset
 - AV_Pathway event_type*= 2
- referral to secondary care appointment
 - source: HES via the RtD derived dataset
 - linked using: AV_Pathway derived dataset
 - AV_Pathway event_type*= 29
- 'first seen' in secondary care
 - source: HES via the RtD derived dataset
 - linked using: AV_Pathway derived dataset
 - AV_Pathway event_type*= 28¹

* AV_pathway event types are identified using numerical coding

- diagnostic imaging
 - source: DID
 - linked using: DID dataset
 - identified using relevant SNOMED-CT and NICIP codes (full details in Appendix 2)
- diagnostic procedure
 - source: HES (APC)
 - linked using: HES dataset
 - identified using relevant OPCS-4 codes (full details in Appendix 2)

A patient may have some, all or none of the events listed above. The first event is determined by selecting the earliest occurring of these events in the 6 months preceding diagnosis, and the corresponding date is established as the start point of the SCDI. In a small number of cases where none of the events are listed the SCDI cannot be calculated (see appendix 1 for more information).

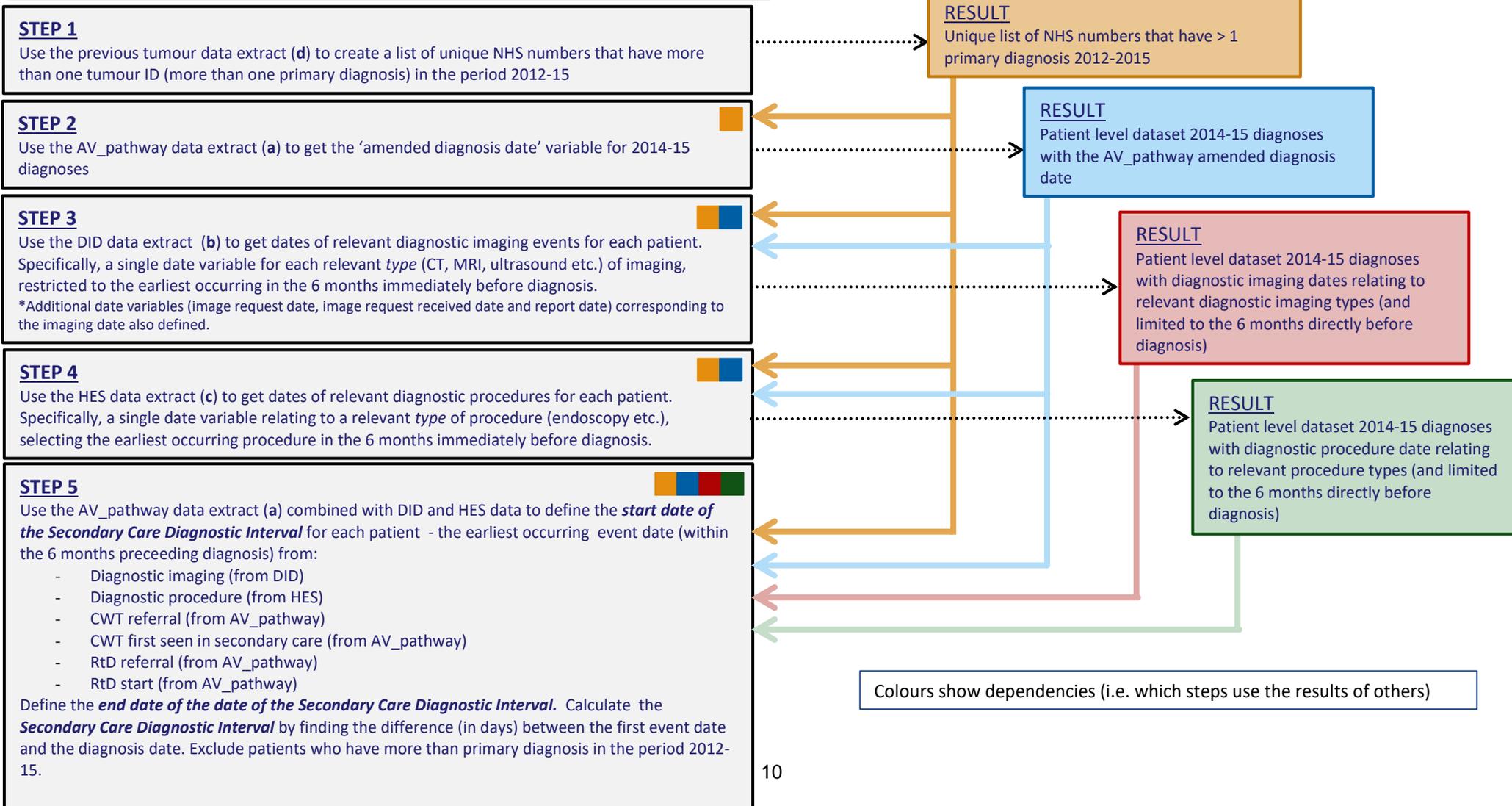
Calculation of interval

The SCDI is calculated by finding the difference in days between the start point (first event date) and the end point (diagnosis date or amended diagnosis date).

Figure 2 – flow chart of process for generating the SCDI (using R software)

SQL data exports:

- a) AV_pathway data (from av_tumour and av_pathway linked datasets)
- b) Diagnostic imaging data (from DID)
- c) Diagnostic procedure data (from HES)
- d) Previous tumour diagnosis data (from av_tumour)



Appendix 1: cancer sites

Broad group	Cancer site	ICD10 codes	% of cohort with a SCDI
Urological	Bladder	C67	98.5
	Kidney	C64-C66	93.3
	Prostate	C61	95.9
Gynaecological	Breast	C50	97.4
	Cervical	C53	97.6
	Ovarian	C56-C57	98.5
	Uterine	C54-C55	98.5
	Vulva/Vagina	C51-C52	94.1
Head and Neck	Colorectal	C18-C20	98.2
	Larynx	C32	98.3
	Oral	C01-C06	94.1
Haematological	Pharynx	C10-C11, C13	97.0
	Leukaemia: acute lymphoblastic	C910	97.2
	Leukaemia: acute myeloid	C920, C924-C926, C928, C930, C940, C942	94.8
	Leukaemia: chronic lymphocytic	C911	89.3
	Leukaemia: chronic myeloid	C921	93.2
	Lymphoma: Hodgkin	C81	98.1
	Lymphoma: non-Hodgkin	C82-C86	95.6
Upper gastro-Intestinal	Multiple myeloma	C90	96.3
	Liver	C22	98.0
	Oesophagus	C15	98.6
	Pancreas	C25	98.7
	Stomach	C16	98.2
	Lung	C33-C34	99.1
	Melanoma	C43	95.2

Appendix 2: diagnostic test codes by cancer site

Process for defining relevant diagnostic tests:

As a first step the CRUK website was consulted to identify a comprehensive list of diagnostic tests for each cancer site. Next, HES (APC) and DID were identified as sources of diagnostic procedures and imaging respectively. Extracts of site cohorts (diagnosis years 2014-15) linked to DID and HES (APC) datasets were used to:

- ascertain which of the tests identified on the CRUK website were available in the data
- consolidate the information from the CRUK website by identifying all tests (in the 6 months preceding diagnosis) occurring in $\geq 1\%$ of patients in each site cohort

Lastly, these site-specific diagnostic test lists were refined in consultation with clinicians to ensure only relevant tests were included.

Diagnostic tests by site:

NB: for each cancer site, relevant diagnostic tests can include either imaging from DID, procedures from HES, or both.

The analysis associated with this SOP used is:

- NICIP and SNOMED-CT coding for diagnostic imaging recorded in DID - exam lookup table V16 [available here](#)
- OPCS-4 coding V4.7 for diagnostic procedures recorded in HES - look-up list and table of equivalence [available here](#)

Bladder

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
CT	CABDO	169070004	CT Abdomen
	CABDOC	32962002	CT Abdomen with contrast
	CABPE	419394006	CT Abdomen and pelvis
	CABPEC	432370003	CT Abdomen and pelvis with contrast
	CCABD	418891003	CT Thorax and abdomen
	CCADBC	429864007	CT Thorax and abdomen with contrast
	CCHAP	418023006	CT Thorax abdomen and pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast
	CCHES	169069000	CT Thorax

	CCHESC	75385009	CT Thorax with contrast
	CKIDB	55501000	CT Renal Both
	CKIDBC	429931008	CT Renal with contrast Both
	CURIT	419084009	CT Urinary tract
	CURITC	440494008	CT Urinary tract with contrast
Ultrasound	UKIDB	306005	US Kidney Both
	UKIDL	306005	US Kidney Lt
	UKIDR	306005	US Kidney Rt
	UUTR	303917008	US Urinary tract
X-Ray	XIVU	32265006	XR Intravenous urogram

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Endoscopy	M421	Endoscopic resection of lesion of bladder
	M422	Endoscopic cauterisation of lesion of bladder
	M451	Diagnostic endoscopic examination of bladder and biopsy of lesion of bladder NEC
	M452	Diagnostic endoscopic examination of bladder and biopsy of lesion of prostate NEC
	M453	Diagnostic endoscopic examination of bladder and biopsy of lesion of bladder using rigid cystoscope
	M454	Diagnostic endoscopic examination of bladder and biopsy of lesion of prostate using rigid cystoscope
	M455	Diagnostic endoscopic examination of bladder using rigid cystoscope
	M458	Other specified diagnostic endoscopic examination of bladder
	M459	Unspecified diagnostic endoscopic examination of bladder

Breast*Diagnostic imaging (DID)*

Image type	NICIP code	SNOMED-CT code	Description
Mammogram	XMAMB	71651007	XR Mammogram Both
	XMAML	71651007	XR Mammogram Lt
	XMAMR	71651007	XR Mammogram Rt
Ultrasound	UMAMB	47079000	US Breast Both
	UMAML	47079000	US Breast Lt
	UMAMR	47079000	US Breast Rt

Cervix

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
CT	CABPE	419394006	CT Abdomen and pelvis
	CABPEC	432370003	CT Abdomen and pelvis with contrast
	CCHAP	418023006	CT Thorax abdomen and pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast
MRI	MABDO	241621009	MRI Abdomen
	MPEGY	826591000000107	MRI Pelvis gynaecological
	MPELV	2690005	MRI Pelvis
	MPELVC	433138001	MRI Pelvis with contrast
Ultrasound	UABDO	45036003	US Abdomen
	UABPE	418394000	US Abdomen and pelvis
	UPELD	826601000000101	US Pelvis TA and transvaginal
	UPELV	24848001	US Pelvis
	UPETV	429915000	US Pelvis transvaginal

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Colposcopy	Q554	Colposcopy of cervix

Colorectal

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED CT code	Description
CT Colonography	CCOLO	425666002	CT Colon
	CCOLOC	445954002	CT Colon (contrast)
	CVCOY	418714002	CT Colonoscopy virtual
	CVCLD	1051291000000100	CT Colonography low dose

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Colonoscopy	G79	Therapeutic endoscopic operations on ileum
	G80	Diagnostic endoscopic examination of ileum
	H20	Endoscopic extirpation of lesion of colon
	H21	Other therapeutic endoscopic operations on colon
	H22	Diagnostic endoscopic examination of colon
	H23	Endoscopic extirpation of lesion of lower bowel using fiberoptic sigmoidoscope
	H24	Other therapeutic endoscopic operations on lower bowel using fiberoptic

H25	sigmoidoscope Diagnostic endoscopic examination of lower bowel using fiberoptic sigmoidoscope
-----	--

Kidney

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
CT	CKIDB	55501000	CT Renal Both
	CKIDBC	429931008	CT Renal with contrast Both
	CURIT	419084009	CT Urinary tract
	CURITC	440494008	CT Urinary tract with contrast
Ultrasound	UKIDB	306005	US Kidney Both
	UKIDL	306005	US Kidney Lt
	UKIDR	306005	US Kidney Rt

Larynx

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
CT	CLARY	241536002	CT Larynx
	CLARYC	431997003	CT Larynx with contrast

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Endoscopy	E371	Diagnostic microendoscopic examination of larynx and biopsy of lesion of larynx
	E361	Diagnostic endoscopic examination of larynx and biopsy of lesion of larynx
	E369	Unspecified diagnostic endoscopic examination of larynx

Leukaemia: acute lymphoblastic

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
CT	CCHAP	418023006	CT Thorax abdomen and pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast
	CCHES	169069000	CT Thorax
	CCHESC	75385009	CT Thorax with contrast
	CNCAP	418332004	CT Neck thorax abdomen and pelvis
	CNCAPC	434438003	CT Neck thorax abdomen pelvis with contrast

Ultrasound	UABDO	45036003	US Abdomen
	UABPE	418394000	US Abdomen and pelvis

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Diagnostic puncture of bone	W365	Diagnostic extraction of bone marrow NEC

Leukaemia: acute myeloid*Diagnostic imaging (DID)*

Image type	NICIP code	SNOMED-CT code	Description
CT	CCABD	418891003	CT Thorax and abdomen
	CCABDC	429864007	CT Thorax and abdomen with contrast
	CCHAP	418023006	CT Thorax abdomen and pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast
	CCHES	169069000	CT Thorax
	CCHESC	75385009	CT Thorax with contrast
	CHRC	315941000000105	CT Chest high resolution
	CHRCHC	75385009	CT Chest high resolution with contrast
	CNCAP	418332004	CT Neck thorax abdomen and pelvis
	CNCAPC	434438003	CT Neck thorax abdomen pelvis with contrast
Ultrasound	UABDO	45036003	US Abdomen
	UABPE	418394000	US Abdomen and pelvis

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Diagnostic puncture of bone	W365	Diagnostic extraction of bone marrow NEC

Leukaemia: chronic lymphocytic*Diagnostic imaging (DID)*

Image type	NICIP code	SNOMED-CT code	Description
CT	CCABD	418891003	CT Thorax and abdomen
	CCABDC	429864007	CT Thorax and abdomen with contrast
	CCHAP	418023006	CT Thorax abdomen and pelvis

	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast
	CCHES	169069000	CT Thorax
	CCHESC	75385009	CT Thorax with contrast
	CHRC	315941000000105	CT Chest high resolution
	CHRCHC	75385009	CT Chest high resolution with contrast
	CNCAP	418332004	CT Neck thorax abdomen and pelvis
	CNCAPC	434438003	CT Neck thorax abdomen pelvis with contrast
Ultrasound	UABDO	45036003	US Abdomen
	UABPE	418394000	US Abdomen and pelvis

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Excision or biopsy of lymph node	T871	Excision or biopsy of scalene lymph node
	T872	Excision or biopsy of cervical lymph node NEC
	T873	Excision or biopsy of axillary lymph node
	T875	Excision or biopsy of para-aortic lymph node
	T877	Excision or biopsy of inguinal lymph node
	T878	Other specified excision or biopsy of lymph node
	T879	Unspecified excision or biopsy of lymph node

Leukaemia: chronic myeloid*Diagnostic imaging (DID)*

Image type	NICIP code	SNOMED-CT code	Description
CT	CCHAP	418023006	CT Thorax abdomen and pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast
	CCHES	169069000	CT Thorax
	CCHESC	75385009	CT Thorax with contrast
	CNCAP	418332004	CT Neck thorax abdomen and pelvis
	CNCAPC	434438003	CT Neck thorax abdomen pelvis with contrast
Ultrasound	UABDO	45036003	US Abdomen
	UABPE	418394000	US Abdomen and pelvis

Liver*Diagnostic imaging (DID)*

Image type	NICIP code	SNOMED-CT code	Description
CT	CABDO	169070004	CT Abdomen
	CABDOC	32962002	CT Abdomen with contrast
	CABPE	419394006	CT Abdomen and pelvis
	CABPEC	432370003	CT Abdomen and pelvis with contrast
	CCABDC	429864007	CT Thorax and abdomen with contrast

	CCHAP	418023006	CT Thorax abdomen and pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast
	CCHES	169069000	CT Thorax
	CCHESC	75385009	CT Thorax with contrast
	CLIVE	241549007	CT Liver
	CLIVEC	429862006	CT Liver with contrast
	CLIVT	438591004	CT Liver triple phase
	CPANSC	429873004	CT Pancreas with contrast
	CPANS	241551006	CT Pancreas
	CPELVC	74710000	CT Pelvis with contrast
MRI	MABDO	241621009	MRI Abdomen
	MDIFF	440408002	MRI Diffusion weighted
	MLIVE	432551009	MRI Liver and spleen
	MLIVEC	431839003	MRI Liver with contrast
	MLIVHC	432633002	MRI Liver with hepatobiliary contrast
	MRLIV	241622002	MRI Liver
Ultrasound	UABDO	45036003	US Abdomen
	UABPE	418394000	US Abdomen and pelvis
	ULIVEB	438300009	US Guided biopsy liver
	ULIVE	105377009	US Liver
	UUPPA	418398002	US Upper abdomen

Lung

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
X-Ray	XCHAB	420233006	XR Chest and abdomen
	XCHES	399208008	XR Chest
	XTHIN	168600009	XR Thoracic inlet
CT	CAPAP	448766000	CT Angiogram pulmonary abdomen pelvis arteries
	CAPUG	419225001	CT Angiogram pulmonary
	CBNTA	440331001	CT Brain neck thorax abdomen pelvis
	CCABD	418891003	CT Thorax and abdomen
	CCABDC	429864007	CT Thorax and abdomen (contrast)
	CCHAP	418023006	CT Thorax abdomen pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis (contrast)
	CCHES	169069000	CT Thorax
	CCHESC	75385009	CT Thorax (contrast)
	CHNTA	711278009	CT Head neck thorax abdomen
	CHNTAC	448760006	CT Head neck thorax abdomen (contrast)
	CHNTAP	440331001	CT Head neck thorax abdomen pelvis
	CHNTPC	444630003	CT Head neck thorax abdomen pelvis (contrast)
	CHRC	315941000000105	CT Chest high resolution
	CHRCHC	75385009	CT Chest high resolution (contrast)
	CHTA	444633001	CT Head thorax and abdomen
	CHTAP	445583005	CT Head thorax abdomen pelvis
	CHTAPC	444674001	CT Head thorax abdomen pelvis (contrast)

CHTH	444708006	CT Head and thorax
CHTHAC	444758004	CT Head thorax abdomen (contrast)
CHTHC	444709003	CT Head and thorax (contrast)
CLDTH	713548006	CT Low dose thorax
CNCAC	433270008	CT Neck thorax abdomen (contrast)
CNCAP	418332004	CT Neck thorax abdomen pelvis
CNCAPC	434438003	CT Neck thorax abdomen pelvis (contrast)
CNCHA	430439002	CT Neck thorax abdomen
CNECH	430448007	CT Neck and thorax
CNECHC	429927002	CT Neck and thorax (contrast)

Lymphoma: Hodgkin

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
CT	CCABD	418891003	CT Thorax and abdomen
	CCABDC	429864007	CT Thorax and abdomen with contrast
	CCHAP	418023006	CT Thorax abdomen and pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast
	CCHES	169069000	CT Thorax
	CCHESC	75385009	CT Thorax with contrast
	CHRC	315941000000105	CT Chest high resolution
	CHRCHC	75385009	CT Chest high resolution with contrast
	CNCAC	433270008	CT Neck thorax and abdomen with contrast
	CNCAP	418332004	CT Neck thorax abdomen and pelvis
	CNCAPC	434438003	CT Neck thorax abdomen pelvis with contrast
	CNCHA	430439002	CT Neck thorax and abdomen
	CNECH	430448007	CT Neck and thorax
	CNECHC	429927002	CT Neck and thorax with contrast
Ultrasound	UABDO	45036003	US Abdomen
	UABPE	418394000	US Abdomen and pelvis
	ULIVE	105377009	US Liver

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Excision or biopsy of lymph node	T871	Excision or biopsy of scalene lymph node
	T872	Excision or biopsy of cervical lymph node NEC
	T873	Excision or biopsy of axillary lymph node
	T874	Excision or biopsy of mediastinal lymph node
	T875	Excision or biopsy of para-aortic lymph node
	T876	Excision or biopsy of porta hepatis lymph node
	T877	Excision or biopsy of inguinal lymph node
	T878	Other specified excision or biopsy of lymph node
	T879	Unspecified excision or biopsy of lymph node

Lymphoma: non-Hodgkin

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
CT	CCABD	418891003	CT Thorax and abdomen
	CCABDC	429864007	CT Thorax and abdomen with contrast
	CCHAP	418023006	CT Thorax abdomen and pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast
	CCHES	169069000	CT Thorax
	CCHESC	75385009	CT Thorax with contrast
	CHRC	315941000000105	CT Chest high resolution
	CHRCHC	75385009	CT Chest high resolution with contrast
	CLIVE	241549007	CT Liver
	CLIVEC	429862006	CT Liver with contrast
	CNCAC	433270008	CT Neck thorax and abdomen with contrast
	CNCAP	418332004	CT Neck thorax abdomen and pelvis
	CNCAPC	434438003	CT Neck thorax abdomen pelvis with contrast
	CNCHA	430439002	CT Neck thorax and abdomen
	CNECH	430448007	CT Neck and thorax
	CNECHC	429927002	CT Neck and thorax with contrast
Ultrasound	UABDO	45036003	US Abdomen
	UABPE	418394000	US Abdomen and pelvis
	ULIVE	105377009	US Liver

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Excision or biopsy of lymph node	T871	Excision or biopsy of scalene lymph node
	T872	Excision or biopsy of cervical lymph node NEC
	T873	Excision or biopsy of axillary lymph node
	T874	Excision or biopsy of mediastinal lymph node
	T875	Excision or biopsy of para-aortic lymph node
	T876	Excision or biopsy of porta hepatis lymph node
	T877	Excision or biopsy of inguinal lymph node
	T878	Other specified excision or biopsy of lymph node
	T879	Unspecified excision or biopsy of lymph node

Melanoma

None

Multiple myeloma

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
CT	CCABD	418891003	CT Thorax and abdomen
	CCABDC	429864007	CT Thorax and abdomen with contrast
	CCHAP	418023006	CT Thorax abdomen and pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast
	CCHES	169069000	CT Thorax
	CCHESC	75385009	CT Thorax with contrast
	CHRC	315941000000105	CT Chest high resolution
	CHRCHC	75385009	CT Chest high resolution with contrast
	CNCAP	418332004	CT Neck thorax abdomen and pelvis
	CNCAPC	434438003	CT Neck thorax abdomen pelvis with contrast
	Ultrasound	UABDO	45036003
UABPE		418394000	US Abdomen and pelvis

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Diagnostic puncture of bone	W365	Diagnostic extraction of bone marrow NEC

Oesophagus

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Endoscopy	G161	Diagnostic fibreoptic endoscopic examination of oesophagus and biopsy of lesion of oesophagus
	G169	Unspecified diagnostic fibreoptic endoscopic examination of oesophagus
	G451	Fibreoptic endoscopic examination of upper gastrointestinal tract and biopsy of lesion of upper gastrointestinal tract
	G459	Unspecified diagnostic fibreoptic endoscopic examination of upper gastrointestinal tract

Oral

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Biopsy	Y204	Fine needle aspiration NOC

Ovary

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
CT	CABDOB	419940006	CT Guided biopsy abdomen
	CABDOC	32962002	CT Abdomen with contrast
	CABPE	419394006	CT Abdomen and pelvis
	CABPEC	432370003	CT Abdomen and pelvis with contrast
	CCHAP	418023006	CT Thorax abdomen and pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast
	CCHESC	75385009	CT Thorax with contrast
	CPELVC	74710000	CT Pelvis with contrast
MRI	MABDO	241621009	MRI Abdomen
	MABDOC	432369004	MRI Abdomen with contrast
	MPEGY	826591000000107	MRI Pelvis gynaecological
	MPELV	2690005	MRI Pelvis
	MPELVC	433138001	MRI Pelvis with contrast
Ultrasound	UABDO	45036003	US Abdomen
	UABDOB	446595004	US Guided biopsy abdomen
	UABPE	418394000	US Abdomen and pelvis
	UBIOPB	277667006	US Guided biopsy
	UPELD	826601000000101	US Pelvis TA and transvaginal
	UPELV	24848001	US Pelvis
	UPETV	429915000	US Pelvis transvaginal
	UPELVB	420041003	US Guided biopsy pelvis

Pancreas

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
Cholangiogram	EERCP	386718000	Endoscopic retrograde cholangiopancreatography
	FPTCH	16747000	Percutaneous transhepatic cholangiogram
	MMRCP	314635004	Magnetic resonance cholangiopancreatography
CT	CABDO	169070004	CT Abdomen
	CABDOC	32962002	CT Abdomen with contrast

	CABPE	419394006	CT Abdomen and pelvis
	CABPEC	432370003	CT Abdomen and pelvis with contrast
	CCABDC	429864007	CT Thorax and abdomen with contrast
	CCHAP	418023006	CT Thorax abdomen and pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast
	CCHES	169069000	CT Thorax
	CCHESC	75385009	CT Thorax with contrast
	CLIVEC	429862006	CT Liver with contrast
	CPANS	241551006	CT Pancreas
	CPANSC	429873004	CT Pancreas with contrast
	CPAND	440514007	CT Pancreas dual phase
	CPELVC	74710000	CT Pelvis with contrast
MRI	MABPA	241625000	MRI Pancreas
	MLIVEC	431839003	MRI Liver with contrast
PET	NCGDW	443632004	NM Ga68 DOTATATE whole body PET CT
	NCGENW	443271005	NM Whole body FDG PET CT
	NGENWO	432675001	NM Whole body PET FDG
Ultrasound	EENDU	450556004	Endoscopic US
	EUPAN	426808008	Endoscopic US pancreas
	UABDO	45036003	US Abdomen
	UABPE	418394000	US Abdomen and pelvis
	ULIVE	105377009	US Liver
	ULIVEB	438300009	US Guided biopsy liver
	UPELV	24848001	US Pelvis
	UUPPA	418398002	US Upper abdomen

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Endoscopy	G451	Fibreoptic endoscopic examination of upper gastrointestinal tract and biopsy of lesion of upper gastrointestinal tract
	G459	Unspecified diagnostic fibreoptic endoscopic examination of upper gastrointestinal tract
	J741	Endoscopic ultrasound examination of pancreas and biopsy of lesion of pancreas
	J749	Unspecified endoscopic ultrasound examination of pancreas
	J439	Unspecified diagnostic endoscopic retrograde examination of bile duct and pancreatic duct

Pharynx

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Endoscopy	E251	Diagnostic endoscopic examination of nasopharynx and biopsy of

		lesion of nasopharynx
E259		Unspecified diagnostic endoscopic examination of pharynx
E252		Diagnostic endoscopic examination of pharynx and biopsy of lesion of pharynx NEC

Prostate

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
MRI	MPEST	75691003	MRI Pelvis prostate
	MPLVPC	963511000000109	MRI Pelvis and prostate with contrast
Ultrasound	UTRPT	22034001	US Transrectal prostate
	UTRPTB	431605004	US Guided biopsy prostate transrectal

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Other operations on outlet of male bladder	M702	Perineal needle biopsy of prostate
	M703	Rectal needle biopsy of prostate

Stomach

Diagnostic procedures (HES)

Procedure type	OPCS-4 code	Description
Endoscopy	G451	Fibreoptic endoscopic examination of upper gastrointestinal tract and biopsy of lesion of upper gastrointestinal tract
	G459	Unspecified diagnostic fibreoptic endoscopic examination of upper gastrointestinal tract
	G436	Fibreoptic endoscopic injection therapy to lesion of upper gastrointestinal tract NEC

Uterus

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
CT	CABPE	419394006	CT Abdomen and pelvis
	CABPEC	432370003	CT Abdomen and pelvis with contrast
	CCHAP	418023006	CT Thorax abdomen and pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast

MRI	MABDO	241621009	MRI Abdomen
	MPEGY	826591000000107	MRI Pelvis gynaecological
	MPELV	2690005	MRI Pelvis
	MPELVC	433138001	MRI Pelvis with contrast
Ultrasound	UABDO	45036003	US Abdomen
	UABPE	418394000	US Abdomen and pelvis
	UPELD	826601000000101	US Pelvis TA and transvaginal
	UPELV	24848001	US Pelvis
	UPETV	429915000	US Pelvis transvaginal

Vagina and vulva

Diagnostic imaging (DID)

Image type	NICIP code	SNOMED-CT code	Description
CT	CABPE	419394006	CT Abdomen and pelvis
	CABPEC	432370003	CT Abdomen and pelvis with contrast
	CCHAP	418023006	CT Thorax abdomen and pelvis
	CCHAPC	433761009	CT Thorax abdomen pelvis with contrast
MRI	MPEGY	826591000000107	MRI Pelvis gynaecological
	MPELV	2690005	MRI Pelvis
	MPELVC	433138001	MRI Pelvis with contrast
Ultrasound	UABDO	45036003	US Abdomen
	UPELD	826601000000101	US Pelvis TA and transvaginal
	UPELV	24848001	US Pelvis
	UPETV	429915000	US Pelvis transvaginal

Appendix 3: example code

This example code outputs patient-level records for those diagnosed with pancreatic cancer (C25). For other sites the ICD10 code and diagnostic test codes will need changing.

(1)SQL exports

- a. AV_pathway data
- b. DID data
- c. HES data
- d. Removing multiple primary tumours data

A. AV_pathway data

```
SELECT t.tumourid, t.diagnosisdatebest, t.NHSNUMBER, t.site_ICD10_02, t.dco,
p.*,
r.route_code, r.final_route,

FROM av2016.av_tumour@casref01 t
LEFT JOIN ANALYSISNCR.AT_PATHWAY@cas1901 p ON p.tumourid=t.tumourid
LEFT JOIN av2016.rtd2016@casref01 r ON r.tumourid = t.tumourid

WHERE t.cascade_inci_flag=1
AND SUBSTR(t.LSOA11_CODE, 1, 1) = 'E'
AND t.diagnosisyear in (2014, 2015)
AND (substr(t.site_ICD10_02,1,3) in ('C25'))
AND t.STATUSOFREGISTRATION = 'F'
AND p.event_type in (1,2,28,29,100);
```

B. DID data

```
SELECT a.tumourid, a.NHSNUMBER, a.diagnosisdatebest,
c.IMAGING_CODE_NICIP, c.IMAGING_CODE_SNOMED_CT, c.PROCEDURE_DATE_CANCER_IMAGING, c.DIAG_TEST_REQ_DATE, c.DIAG_TEST_REQ_REC_DATE,
c.SERVICE_REPORT_ISSUE_DATE
```

CAS-SOP: Defining the Secondary Care Diagnostic Interval

```
FROM av2016.av_tumour@casref01 a
LEFT JOIN DIDS.DIDS_2013_TO_2015 c ON a.nhsnumber = c.nhs_number

WHERE a.diagnosisyear in ('2014', '2015')
      AND a.SITE_ICD10_O2_3CHAR in ('C25')
      AND a.LSOA11_CODE LIKE 'E%'
      AND a.STATUSOFREGISTRATION = 'F'
      AND cascade_inci_flag=1;
```

C. HES (APC) data

```
SELECT a.patientid, a.tumourid, a.diagnosisdatebest, a.diagnosisyear, a.birthdatebest,
a.NHSNUMBER, a.SITE_ICD10_O2_3CHAR,
```

```
(SELECT apc_oper.OPERTN FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=1) as OPERTN_01,
(SELECT apc_oper.OPERTN FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=2) as OPERTN_02,
(SELECT apc_oper.OPERTN FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=3) as OPERTN_03,
(SELECT apc_oper.OPERTN FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=4) as OPERTN_04,
(SELECT apc_oper.OPERTN FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=5) as OPERTN_05,
(SELECT apc_oper.OPERTN FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=6) as OPERTN_06,
(SELECT apc_oper.OPERTN FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=7) as OPERTN_07,
(SELECT apc_oper.OPERTN FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=8) as OPERTN_08,
(SELECT apc_oper.OPERTN FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=9) as OPERTN_09,
(SELECT apc_oper.OPERTN FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=10) as OPERTN_10,
(SELECT apc_oper.OPERTN FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=11) as OPERTN_11,
(SELECT apc_oper.OPERTN FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=12) as OPERTN_12,
(SELECT apc_oper.OPERTN FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=13) as OPERTN_13,
```


CAS-SOP: Defining the Secondary Care Diagnostic Interval

```
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=10) as OPDATE_10,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=11) as OPDATE_11,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=12) as OPDATE_12,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=13) as OPDATE_13,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=14) as OPDATE_14,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=15) as OPDATE_15,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=16) as OPDATE_16,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=17) as OPDATE_17,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=18) as OPDATE_18,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=19) as OPDATE_19,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=20) as OPDATE_20,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=21) as OPDATE_21,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=22) as OPDATE_22,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=23) as OPDATE_23,
(SELECT apc_oper.OPDATE FROM HES2017.HESAPC_OPERTN apc_oper WHERE apc.DATAYEAR = apc_oper.DATAYEAR AND apc.EPIKEYANON = apc_oper.EPIKEYANON AND
apc_oper.POS=24) as OPDATE_24
```

```
FROM av2016.av_tumour a
```

```
LEFT JOIN hes2017.hes_linkage_AV_APC li on a.patientid = li.patientid
```

```
LEFT JOIN hes2017.hesapc apc on li.datayear = apc.datayear and li.epikeyanon = apc.epikeyanon
```

```
WHERE a.diagnosisyear in ('2014', '2015')
      AND a.SITE_ICD10_O2_3CHAR in ('C25')
      AND a.LSOA11_CODE LIKE 'E%'
      AND a.STATUSOFREGISTRATION = 'F'
      AND a.cascade_inci_flag=1;
```

D. Removing multiple primary tumour data

```
SELECT tumourid, NHSNUMBER, diagnosisyear
FROM AV2016.AV_TUMOUR
WHERE diagnosisyear in ('2012', '2013', '2014', '2015')
      AND SITE_ICD10_02_3CHAR in ('C25')
      AND LSOA11_CODE LIKE 'E%'
      AND STATUSOFREGISTRATION = 'F'
      AND cascade_inci_flag=1;
```

(2) R code

```
#----- PACKAGES -----  
  
#install.packages("plyr ")  
#install.packages("dplyr")  
#install.packages("zoo")  
library(plyr)  
library(dplyr)  
library("zoo")  
  
#----- IMPORTING DATA -----  
  
#### DATASETS ####  
  
# (1) AV_PATHWAY  
# (2) DID  
# (3) HES  
# (4) MULTIPLE PRIMARIES  
  
av_pathway <- read.csv("filepath/your_data1.csv",header=TRUE, sep=",")  
DID <- read.csv("filepath/your_data2.csv",header=TRUE, sep=",")  
HES <- read.csv("filepath/your_data3.csv",header=TRUE, sep=",")  
mulprims <- read.csv("filepath/your_data4.csv",header=TRUE, sep=",")  
  
#----- MULTIPLE PRIMARIES -----  
  
#----- Generating Multiple Primaries Flag -----  
  
# this gives with missing (na) values  
dupNHS <- mulprims %>% group_by(NHSNUMBER) %>% filter(n())>1)  
dupNHS <- subset(dupNHS, DIAGNOSISYEAR== 2014 | DIAGNOSISYEAR==2015, c("NHSNUMBER"))  
dupNHS <- data.frame(unique(dupNHS$NHSNUMBER))  
dupNHS$flag <- 1  
names(dupNHS)[names(dupNHS)=="unique.dupNHS.NHSNUMBER."] <-"NHSNUMBER"  
  
#----- DERIVED DIAGNOSIS DATE – AV_PATH -----  
  
deriveddiagdate <- subset(av_pathway, event_type == 100)
```

CAS-SOP: Defining the Secondary Care Diagnostic Interval

#----- Dropping Multiple Primaries -----

merging duplicate NHS number flag

```
deriveddiagdate <- merge(x = deriveddiagdate, y = dupNHS, by = "NHSNUMBER", all.x = TRUE)
deriveddiagdate$flag[is.na(deriveddiagdate$flag)] <- 0
```

dropping those with a dupNHS flag=1

```
deriveddiagdate <- deriveddiagdate[ which(deriveddiagdate$flag!=1), ]
```

#----- Generating New Diagnosis Date -----

AV_PATH event data (derived diagnosis date)

NOTE the below code requires 'zoo' package

```
deriveddiagdate$DerDiagDate <- as.Date(deriveddiagdate$EVENT_DATE, format = "%d-%b-%y")
```

```
deriveddiagdate <- deriveddiagdate[c("NHSNUMBER", "DerDiagDate")]
```

#----- DID -----

#----- Dropping Multiple Primaries -----

merging duplicate NHS number flag

```
DID <- merge(x = DID, y = dupNHS, by = "NHSNUMBER", all.x = TRUE)
DID$flag[is.na(DID$flag)] <- 0
```

dropping those with a dupNHS flag=1

```
DID <- DID[ which(DID$flag!=1), ]
```

#----- Reformatting Dates -----

```
DID$diagdate <- as.Date(DID$DIAGNOSISDATEBEST, format = "%d-%b-%y")
DID$didddate <- as.Date(DID$PROCEDURE_DATE_CANCER_IMAGING, format = "%d-%b-%y")
DID$testreqdate <- as.Date(DID$DIAG_TEST_REQ_DATE, format = "%d-%b-%y")
DID$testreqrecdate <- as.Date(DID$DIAG_TEST_REQ_REC_DATE, format = "%d-%b-%y")
DID$testreport <- as.Date(DID$SERVICE_REPORT_ISSUE_DATE, format = "%d-%b-%y")
```

#----- Reformatting Non-Date Variables -----

SNOMED

CAS-SOP: Defining the Secondary Care Diagnostic Interval

```
DID$SNOMED_final <- as.character(DID$IMAGING_CODE_SNOMED_CT)
```

```
# NICIP
```

```
DID$NICIP_final <- DID$IMAGING_CODE_NICIP
```

```
DID$NICIP_final[DID$NICIP_final==""] <- NA
```

```
#----- Merging Alternative Diagnosis Date -----
```

```
# merging duplicate NHS number flag
```

```
DID <- merge(x = DID, y = deriveddiagdate, by = "NHSNUMBER", all.x = TRUE)
```

```
# generating new variable
```

```
DID$diagdiff <- DID$diagdate - DID$DerDiagDate
```

```
DID$diagdate_new <- as.Date(ifelse((DID$diagdiff >0 & DID$diagdiff <=28) & !is.na(DID$diagdiff), DID$DerDiagDate, DID$diagdate))
```

```
#----- Generating Ultrasound Flag -----
```

```
DID$US_any[DID$NICIP_final=="UABDO" | DID$SNOMED_final=="45036003" |  
  DID$NICIP_final=="ULIVEB" | DID$SNOMED_final=="438300009" |  
  DID$NICIP_final=="UABPE" | DID$SNOMED_final=="418394000" |  
  DID$NICIP_final=="EENDU" | DID$SNOMED_final=="450556004" |  
  DID$NICIP_final=="UUPPA" | DID$SNOMED_final=="418398002" |  
  DID$NICIP_final=="ULIVE" | DID$SNOMED_final=="105377009" |  
  DID$NICIP_final=="EUPAN" | DID$SNOMED_final=="426808008" |  
  DID$NICIP_final=="UPELV" | DID$SNOMED_final=="24848001"] <- 1
```

```
DID$US_any[is.na(DID$US_any)] <- 0
```

```
#----- Generating CT Flag -----
```

```
DID$CT_any[DID$NICIP_final=="CCHAPC" | DID$SNOMED_final=="433761009" |  
  DID$NICIP_final=="CABPEC" | DID$SNOMED_final=="432370003" |  
  DID$NICIP_final=="CPANSC" | DID$SNOMED_final=="429873004" |  
  DID$NICIP_final=="CCHESC" | DID$SNOMED_final=="75385009" |  
  DID$NICIP_final=="CABDOC" | DID$SNOMED_final=="32962002" |  
  DID$NICIP_final=="CCHES" | DID$SNOMED_final=="169069000" |  
  DID$NICIP_final=="CCHAP" | DID$SNOMED_final=="418023006" |  
  DID$NICIP_final=="CPANS" | DID$SNOMED_final=="241551006" |  
  DID$NICIP_final=="CABPE" | DID$SNOMED_final=="419394006" |  
  DID$NICIP_final=="CABDO" | DID$SNOMED_final=="169070004" |  
  DID$NICIP_final=="CCABDC" | DID$SNOMED_final=="429864007"]
```

CAS-SOP: Defining the Secondary Care Diagnostic Interval

```
DID$NICIP_final=="CPELVC" | DID$SNOMED_final=="74710000" |  
DID$NICIP_final=="CPAND" | DID$SNOMED_final=="440514007" |  
DID$NICIP_final=="CLIVEC" | DID$SNOMED_final=="429862006" ] <- 1
```

```
DID$CT_any[is.na(DID$CT_any)] <- 0
```

```
#----- Generating Cholangiogram Flag -----
```

```
DID$CHOL_any[DID$NICIP_final=="EERCP" | DID$SNOMED_final=="386718000" |  
DID$NICIP_final=="MMRCP" | DID$SNOMED_final=="314635004" |  
DID$NICIP_final=="FPTCH" | DID$SNOMED_final=="16747000" ] <- 1
```

```
DID$CHOL_any[is.na(DID$CHOL_any)] <- 0
```

```
#----- Generating PET Flag -----
```

```
DID$PET_any[DID$NICIP_final=="NCGENW" | DID$SNOMED_final=="443271005" |  
DID$NICIP_final=="NGENWO" | DID$SNOMED_final=="432675001" |  
DID$NICIP_final=="NCGDW" | DID$SNOMED_final=="443632004" ] <- 1
```

```
DID$PET_any[is.na(DID$PET_any)] <- 0
```

```
#----- Generating MRI Flag -----
```

```
DID$MRI_any[DID$NICIP_final=="MLIVEC" | DID$SNOMED_final=="431839003" |  
DID$NICIP_final=="MABPA" | DID$SNOMED_final=="241625000" ] <- 1
```

```
DID$MRI_any[is.na(DID$MRI_any)] <- 0
```

```
#----- Generating Relevant 6 Month Imaging -----
```

```
DID$imaging_diff <- DID$diagdate_new - DID$didddate
```

```
DID$US6mos <- as.Date(ifelse((DID$US_any==1 & DID$imaging_diff<= 182 & DID$imaging_diff >=0), DID$didddate, NA))  
DID$CT6mos <- as.Date(ifelse((DID$CT_any==1 & DID$imaging_diff<= 182 & DID$imaging_diff >=0), DID$didddate, NA))  
DID$CHOL6mos <- as.Date(ifelse((DID$CHOL_any==1 & DID$imaging_diff<= 182 & DID$imaging_diff >=0), DID$didddate, NA))  
DID$PET6mos <- as.Date(ifelse((DID$PET_any==1 & DID$imaging_diff<= 182 & DID$imaging_diff >=0), DID$didddate, NA))  
DID$MRI6mos <- as.Date(ifelse((DID$MRI_any==1 & DID$imaging_diff<= 182 & DID$imaging_diff >=0), DID$didddate, NA))
```

CAS-SOP: Defining the Secondary Care Diagnostic Interval

```
DID$US6mos_req <- as.Date(ifelse((DID$US_any==1 & DID$diagdate_new - DID$testreqdate <= 182 & DID$diagdate_new - DID$testreqdate >=0), DID$testreqdate, NA))
DID$CT6mos_req <- as.Date(ifelse((DID$CT_any==1 & DID$diagdate_new - DID$testreqdate <= 182 & DID$diagdate_new - DID$testreqdate >=0), DID$testreqdate, NA))
DID$CHOL6mos_req <- as.Date(ifelse((DID$CHOL_any==1 & DID$diagdate_new - DID$testreqdate <= 182 & DID$diagdate_new - DID$testreqdate >=0), DID$testreqdate, NA))
DID$PET6mos_req <- as.Date(ifelse((DID$PET_any==1 & DID$diagdate_new - DID$testreqdate <= 182 & DID$diagdate_new - DID$testreqdate >=0), DID$testreqdate, NA))
DID$MRI6mos_req <- as.Date(ifelse((DID$MRI_any==1 & DID$diagdate_new - DID$testreqdate <= 182 & DID$diagdate_new - DID$testreqdate >=0), DID$testreqdate, NA))
DID$US6mos_req[is.na(DID$US6mos)] <- NA
DID$CT6mos_req[is.na(DID$CT6mos)] <- NA
DID$CHOL6mos_req[is.na(DID$CHOL6mos)] <- NA
DID$PET6mos_req[is.na(DID$PET6mos)] <- NA
DID$MRI6mos_req[is.na(DID$MRI6mos)] <- NA
```

```
DID$US6mos_rec <- as.Date(ifelse((DID$US_any==1 & DID$diagdate_new - DID$testreqrec <= 182 & DID$diagdate_new - DID$testreqrec >=0), DID$testreqrec, NA))
DID$CT6mos_rec <- as.Date(ifelse((DID$CT_any==1 & DID$diagdate_new - DID$testreqrec <= 182 & DID$diagdate_new - DID$testreqrec >=0), DID$testreqrec, NA))
DID$CHOL6mos_rec <- as.Date(ifelse((DID$CHOL_any==1 & DID$diagdate_new - DID$testreqrec <= 182 & DID$diagdate_new - DID$testreqrec >=0), DID$testreqrec, NA))
DID$PET6mos_rec <- as.Date(ifelse((DID$PET_any==1 & DID$diagdate_new - DID$testreqrec <= 182 & DID$diagdate_new - DID$testreqrec >=0), DID$testreqrec, NA))
DID$MRI6mos_rec <- as.Date(ifelse((DID$MRI_any==1 & DID$diagdate_new - DID$testreqrec <= 182 & DID$diagdate_new - DID$testreqrec >=0), DID$testreqrec, NA))
DID$US6mos_rec[is.na(DID$US6mos)] <- NA
DID$CT6mos_rec[is.na(DID$CT6mos)] <- NA
DID$CHOL6mos_rec[is.na(DID$CHOL6mos)] <- NA
DID$PET6mos_rec[is.na(DID$PET6mos)] <- NA
DID$MRI6mos_rec[is.na(DID$MRI6mos)] <- NA
```

```
DID$US6mos_report <- as.Date(ifelse((DID$US_any==1 & DID$diagdate_new - DID$testreport <= 182 & DID$diagdate_new - DID$testreport >=0), DID$testreport, NA))
DID$CT6mos_report <- as.Date(ifelse((DID$CT_any==1 & DID$diagdate_new - DID$testreport <= 182 & DID$diagdate_new - DID$testreport >=0), DID$testreport, NA))
DID$CHOL6mos_report <- as.Date(ifelse((DID$CHOL_any==1 & DID$diagdate_new - DID$testreport <= 182 & DID$diagdate_new - DID$testreport >=0), DID$testreport, NA))
DID$PET6mos_report <- as.Date(ifelse((DID$PET_any==1 & DID$diagdate_new - DID$testreport <= 182 & DID$diagdate_new - DID$testreport >=0), DID$testreport, NA))
DID$MRI6mos_report <- as.Date(ifelse((DID$MRI_any==1 & DID$diagdate_new - DID$testreport <= 182 & DID$diagdate_new - DID$testreport >=0), DID$testreport, NA))
DID$US6mos_report[is.na(DID$US6mos)] <- NA
DID$CT6mos_report[is.na(DID$CT6mos)] <- NA
DID$CHOL6mos_report[is.na(DID$CHOL6mos)] <- NA
DID$PET6mos_report[is.na(DID$PET6mos)] <- NA
DID$MRI6mos_report[is.na(DID$MRI6mos)] <- NA
```

#----- Collapsing -----

```
DID_collapsed <- DID %>%
  group_by(NHSNUMBER) %>%
  summarize(diagdate_new_DID = as.Date(if(all(is.na(diagdate_new))) NA else min(diagdate_new, na.rm = TRUE)),
            US6mos = as.Date(if(all(is.na(US6mos))) NA else min(US6mos, na.rm = TRUE)),
            CT6mos = as.Date(if(all(is.na(CT6mos))) NA else min(CT6mos, na.rm = TRUE)),
            CHOL6mos = as.Date(if(all(is.na(CHOL6mos))) NA else min(CHOL6mos, na.rm = TRUE)),
            PET6mos = as.Date(if(all(is.na(PET6mos))) NA else min(PET6mos, na.rm = TRUE)),
            MRI6mos = as.Date(if(all(is.na(MRI6mos))) NA else min(MRI6mos, na.rm = TRUE)),
```

CAS-SOP: Defining the Secondary Care Diagnostic Interval

```
US6mos_req = as.Date(if(all(is.na(US6mos_req))) NA else min(US6mos_req, na.rm = TRUE)),
CT6mos_req = as.Date(if(all(is.na(CT6mos_req))) NA else min(CT6mos_req, na.rm = TRUE)),
CHOL6mos_req = as.Date(if(all(is.na(CHOL6mos_req))) NA else min(CHOL6mos_req, na.rm = TRUE)),
PET6mos_req = as.Date(if(all(is.na(PET6mos_req))) NA else min(PET6mos_req, na.rm = TRUE)),
MRI6mos_req = as.Date(if(all(is.na(MRI6mos_req))) NA else min(MRI6mos_req, na.rm = TRUE)),
US6mos_rec = as.Date(if(all(is.na(US6mos_rec))) NA else min(US6mos_rec, na.rm = TRUE)),
CT6mos_rec = as.Date(if(all(is.na(CT6mos_rec))) NA else min(CT6mos_rec, na.rm = TRUE)),
CHOL6mos_rec = as.Date(if(all(is.na(CHOL6mos_rec))) NA else min(CHOL6mos_rec, na.rm = TRUE)),
PET6mos_rec = as.Date(if(all(is.na(PET6mos_rec))) NA else min(PET6mos_rec, na.rm = TRUE)),
MRI6mos_rec = as.Date(if(all(is.na(MRI6mos_rec))) NA else min(MRI6mos_rec, na.rm = TRUE)),
US6mos_report = as.Date(if(all(is.na(US6mos_report))) NA else min(US6mos_report, na.rm = TRUE)),
CT6mos_report = as.Date(if(all(is.na(CT6mos_report))) NA else min(CT6mos_report, na.rm = TRUE)),
CHOL6mos_report = as.Date(if(all(is.na(CHOL6mos_report))) NA else min(CHOL6mos_report, na.rm = TRUE)),
PET6mos_report = as.Date(if(all(is.na(PET6mos_report))) NA else min(PET6mos_report, na.rm = TRUE)),
MRI6mos_report = as.Date(if(all(is.na(MRI6mos_report))) NA else min(MRI6mos_report, na.rm = TRUE))
```

```
#----- HES -----
```

```
#----- Dropping Multiple Primaries -----
```

```
# merging duplicate NHS number flag
```

```
HES <- merge(x = HES, y = dupNHS, by = "NHSNUMBER", all.x = TRUE)
HES$flag[is.na(HES$flag)] <- 0
```

```
# dropping those with a dupNHS flag=1
```

```
HES <- HES[ which(HES$flag!=1), ]
```

```
#----- Reformating Diagnosis Date -----
```

```
HES$diagdate_HES <- as.Date(HES$DIAGNOSISDATEBEST, format = "%d-%b-%y")
```

```
#----- Reformating Opcodes -----
```

```
OPcount <- seq(1, 24, 1)
OPcount_char <- sprintf("%02d", OPcount)
```

```
for (i in OPcount_char) {
  HES[paste("OPDATE_",i,sep="")] <- as.Date(HES[[paste("OPDATE_",i,sep=")]], format = "%d-%b-%y")
}
```

```
#----- Merging Alternative Diagnosis Date -----
```

CAS-SOP: Defining the Secondary Care Diagnostic Interval

merging duplicate NHS number flag

```
HES <- merge(x = HES, y = deriveddiagdate, by = "NHSNUMBER", all.x = TRUE)
```

generating new variable

```
HES$diagdiff <- HES$diagdate_HES - HES$DerDiagDate
```

```
HES$diagdate_new <- as.Date(ifelse((HES$diagdiff > 0 & HES$diagdiff <= 28) & !is.na(HES$diagdiff), HES$DerDiagDate, HES$diagdate))
```

#----- Flagging Relevant Procedure Types -----

```
for (i in OPcount_char) {
```

```
  HES[paste("HESendo_", i, sep="")] <- ifelse(HES[paste("OPERTN_", i, sep="")] == "G451" | HES[paste("OPERTN_", i, sep="")] == "G459" | HES[paste("OPERTN_", i, sep="")] == "J741" |  
HES[paste("OPERTN_", i, sep="")] == "J749" | HES[paste("OPERTN_", i, sep="")] == "J439", 1, 0)  
  HES[paste("HESendo_", i, sep="")] [is.na(HES[paste("HESendo_", i, sep="")])] <- 0
```

```
}
```

#----- Restricting to 6 Month timeframe -----

```
for (i in OPcount_char) {
```

```
  HES[paste("DIFDATEFLAG_", i, sep="")] <- ifelse(HES[[paste("OPDATE_", i, sep="")]] - HES$diagdate_new <= 0 &  
HES[[paste("OPDATE_", i, sep="")]] - HES$diagdate_new >= -182, 1, 0)
```

```
}
```

```
for (i in OPcount_char) {
```

```
  HES[paste("HESendo6m_", i, sep="")] <- as.Date(as.integer(ifelse(HES[[paste("HESendo_", i, sep="")]] == 1 & HES[[paste("DIFDATEFLAG_", i, sep="")]] == 1,  
HES[[paste("OPDATE_", i, sep="")]], NA)))
```

```
}
```

#----- Collapsing -----

```
HES_collapsed <- HES %>%
```

```
  group_by(NHSNUMBER) %>%
```

```
  summarize(HESendo6m_01 = as.Date(if(all(is.na(HESendo6m_01))) NA else min(HESendo6m_01, na.rm = TRUE)),
```

```
    HESendo6m_02 = as.Date(if(all(is.na(HESendo6m_02))) NA else min(HESendo6m_02, na.rm = TRUE)),
```

```
    HESendo6m_03 = as.Date(if(all(is.na(HESendo6m_03))) NA else min(HESendo6m_03, na.rm = TRUE)),
```

```
    HESendo6m_04 = as.Date(if(all(is.na(HESendo6m_04))) NA else min(HESendo6m_04, na.rm = TRUE)),
```

```
    HESendo6m_05 = as.Date(if(all(is.na(HESendo6m_05))) NA else min(HESendo6m_05, na.rm = TRUE)),
```

```
    HESendo6m_06 = as.Date(if(all(is.na(HESendo6m_06))) NA else min(HESendo6m_06, na.rm = TRUE)),
```

```
    HESendo6m_07 = as.Date(if(all(is.na(HESendo6m_07))) NA else min(HESendo6m_07, na.rm = TRUE)),
```

```
    HESendo6m_08 = as.Date(if(all(is.na(HESendo6m_08))) NA else min(HESendo6m_08, na.rm = TRUE)),
```

CAS-SOP: Defining the Secondary Care Diagnostic Interval

```
HESendo6m_09 = as.Date(if(all(is.na(HESendo6m_09))) NA else min(HESendo6m_09, na.rm = TRUE)),
HESendo6m_10 = as.Date(if(all(is.na(HESendo6m_10))) NA else min(HESendo6m_10, na.rm = TRUE)),
HESendo6m_11 = as.Date(if(all(is.na(HESendo6m_11))) NA else min(HESendo6m_11, na.rm = TRUE)),
HESendo6m_12 = as.Date(if(all(is.na(HESendo6m_12))) NA else min(HESendo6m_12, na.rm = TRUE)),
HESendo6m_13 = as.Date(if(all(is.na(HESendo6m_13))) NA else min(HESendo6m_13, na.rm = TRUE)),
HESendo6m_14 = as.Date(if(all(is.na(HESendo6m_14))) NA else min(HESendo6m_14, na.rm = TRUE)),
HESendo6m_15 = as.Date(if(all(is.na(HESendo6m_15))) NA else min(HESendo6m_15, na.rm = TRUE)),
HESendo6m_16 = as.Date(if(all(is.na(HESendo6m_16))) NA else min(HESendo6m_16, na.rm = TRUE)),
HESendo6m_17 = as.Date(if(all(is.na(HESendo6m_17))) NA else min(HESendo6m_17, na.rm = TRUE)),
HESendo6m_18 = as.Date(if(all(is.na(HESendo6m_18))) NA else min(HESendo6m_18, na.rm = TRUE)),
HESendo6m_19 = as.Date(if(all(is.na(HESendo6m_19))) NA else min(HESendo6m_19, na.rm = TRUE)),
HESendo6m_20 = as.Date(if(all(is.na(HESendo6m_20))) NA else min(HESendo6m_20, na.rm = TRUE)),
HESendo6m_21 = as.Date(if(all(is.na(HESendo6m_21))) NA else min(HESendo6m_21, na.rm = TRUE)),
HESendo6m_22 = as.Date(if(all(is.na(HESendo6m_22))) NA else min(HESendo6m_22, na.rm = TRUE)),
HESendo6m_23 = as.Date(if(all(is.na(HESendo6m_23))) NA else min(HESendo6m_23, na.rm = TRUE)),
HESendo6m_24 = as.Date(if(all(is.na(HESendo6m_24))) NA else min(HESendo6m_24, na.rm = TRUE))
```

```
HES_collapsed <- mutate(HES_collapsed,
  HES_ENDOdate6m= pmin(HESendo6m_01, HESendo6m_02, HESendo6m_03, HESendo6m_04, HESendo6m_05, HESendo6m_06,
    HESendo6m_07, HESendo6m_08, HESendo6m_09, HESendo6m_10, HESendo6m_11, HESendo6m_12,
    HESendo6m_13, HESendo6m_14, HESendo6m_15, HESendo6m_16, HESendo6m_17, HESendo6m_18,
    HESendo6m_19, HESendo6m_20, HESendo6m_21, HESendo6m_22, HESendo6m_23, HESendo6m_24, na.rm=TRUE))
```

```
HES_collapsed <- HES_collapsed %>%
  select(NHSNUMBER, HES_ENDOdate6m)
```

```
#----- AV_PATHWAY -----
```

```
#----- Dropping Multiple Primaries -----
```

```
av_pathway <- merge(x = av_pathway, y = dupNHS, by = "NHSNUMBER", all.x = TRUE)
av_pathway$flag[is.na(av_pathway$flag)] <- 0
```

```
av_pathway <- av_pathway[ which(av_pathway$flag!=1), ]
```

```
#----- Reformatting Dates -----
```

```
av_pathway$av_event_date <- as.Date(av_pathway$EVENT_DATE, format = "%d-%b-%y")
av_pathway$diag_date_best <- as.Date(av_pathway$DIAGNOSISDATEBEST, format = "%d-%b-%y")
```

```
#----- Merging Alternative Diagnosis Date -----
```

CAS-SOP: Defining the Secondary Care Diagnostic Interval

merging duplicate NHS number flag

```
av_pathway <- merge(x = av_pathway, y = deriveddiagdate, by = "NHSNUMBER", all.x = TRUE)
```

generating new variable

```
av_pathway$diagdiff <- av_pathway$diag_date_best - av_pathway$DerDiagDate
```

```
av_pathway$diagdate_new <- as.Date(ifelse((av_pathway$diagdiff >0 & av_pathway$diagdiff <=28) & !is.na(av_pathway$diagdiff), av_pathway$DerDiagDate, av_pathway$diag_date_best))
```

#----- Dropping events -----

dropping events outside the 6 month timeframe

```
av_pathway$difdate <- av_pathway$sav_event_date - av_pathway$diagdate_new  
summary(av_pathway$difdate)
```

```
av_pathway <- filter(av_pathway, difdate >=-182 & difdate<=0)
```

```
summary(av_pathway$NHSNUMBER)  
length(unique(av_pathway$NHSNUMBER))
```

#----- Generating AVpath Event Dates -----

```
av_pathway$CWT_ref_date <- as.Date(ifelse(1==av_pathway$EVENT_TYPE, av_pathway$sav_event_date, NA))  
av_pathway$CWT_FS_date <- as.Date(ifelse(2==av_pathway$EVENT_TYPE, av_pathway$sav_event_date, NA))  
av_pathway$RtD_start_HES_date <- as.Date(ifelse(28==av_pathway$EVENT_TYPE, av_pathway$sav_event_date, NA))  
av_pathway$RtD_ref_HES_date <- as.Date(ifelse(29==av_pathway$EVENT_TYPE, av_pathway$sav_event_date, NA))
```

#----- Collapsing -----

```
av_pathway <- arrange(av_pathway, NHSNUMBER, EVENT_TYPE, av_event_date)
```

```
av_pathway_collapsed <- av_pathway %>%  
  group_by(NHSNUMBER) %>%  
  summarize(CWT_ref_date = as.Date(if(all(is.na(CWT_ref_date))) NA else min(CWT_ref_date, na.rm = TRUE)),  
            CWT_FS_date = as.Date(if(all(is.na(CWT_FS_date))) NA else min(CWT_FS_date, na.rm = TRUE)),  
            DID_date = as.Date(if(all(is.na(DID_date))) NA else min(DID_date, na.rm = TRUE)),  
            RtD_start_HES_date = as.Date(if(all(is.na(RtD_start_HES_date))) NA else min(RtD_start_HES_date, na.rm = TRUE)),  
            RtD_ref_HES_date = as.Date(if(all(is.na(RtD_ref_HES_date))) NA else min(RtD_ref_HES_date, na.rm = TRUE)),  
            diagdate_new = as.Date(if(all(is.na(diagdate_new))) NA else min(diagdate_new, na.rm = TRUE)),  
            diag_date_best = as.Date(if(all(is.na(diag_date_best))) NA else min(diag_date_best, na.rm = TRUE)),
```

CAS-SOP: Defining the Secondary Care Diagnostic Interval

```
tumid = TUMOURID[which(!is.na(TUMOURID))[1]],  
dco = DCO[which(!is.na(DCO))[1]],  
final_route = FINAL_ROUTE[which(!is.na(FINAL_ROUTE))[1]],  
route_code = ROUTE_CODE[which(!is.na(ROUTE_CODE))[1]]
```

```
#----- Merging DID -----
```

```
av_pathway_collapsed <- merge(x = av_pathway_collapsed, y = DID_collapsed, by = "NHSNUMBER", all.x = TRUE)
```

```
#----- Merging HES -----
```

```
av_pathway_collapsed <- merge(x = av_pathway_collapsed, y = HES_collapsed, by = "NHSNUMBER", all.x = TRUE)
```

```
#----- Dropping Used Datasets -----
```

```
# remove as appropriate
```

```
rm(mulprims, dupNHS, deriveddiagdate, DID, DID_collapsed, HES, HES_collapsed, OPcount, OPcount_char, i, av_pathway)
```

```
#----- Cleaning 'demographic' variables -----
```

```
# RtD
```

```
table(av_pathway_collapsed$final_route)  
table(is.na(av_pathway_collapsed$final_route))  
class(av_pathway_collapsed$final_route)
```

```
av_pathway_collapsed$final_route[is.na(av_pathway_collapsed$final_route)] <- "Unknown"
```

```
av_pathway_collapsed$final_route_coded <- as.factor(av_pathway_collapsed$final_route)
```

```
table(av_pathway_collapsed$final_route_coded)  
table(is.na(av_pathway_collapsed$final_route_coded))  
class(av_pathway_collapsed$final_route_coded)
```

```
# DCO
```

```
table(av_pathway_collapsed$dco)  
table(is.na(av_pathway_collapsed$dco))  
class(av_pathway_collapsed$dco)
```

```
av_pathway_collapsed$dco <- as.factor(av_pathway_collapsed$dco)
```

CAS-SOP: Defining the Secondary Care Diagnostic Interval

#----- Generating DID Combination Variable -----

```
av_pathway_collapsed <- mutate(av_pathway_collapsed,  
  DID6m= pmin(US6mos, CT6mos, CHOL6mos, PET6mos, MRI6mos, US6mos_req, CT6mos_req, CHOL6mos_req, PET6mos_req, MRI6mos_req, US6mos_rec, CT6mos_rec,  
  CHOL6mos_rec, PET6mos_rec, MRI6mos_rec, na.rm=TRUE))
```

#----- Generating First Event Variable -----

```
av_pathway_collapsed <- mutate(av_pathway_collapsed,  
  FEdate= pmin(CWT_ref_date, CWT_FS_date, RtD_start_HES_date, RtD_ref_HES_date, DID6m, HES_ENDOdate6m, na.rm=TRUE))
```

```
table(is.na(av_pathway_collapsed$FEdate))  
length(unique(av_pathway_collapsed$FEdate))  
class(av_pathway_collapsed$FEdate)  
summary(av_pathway_collapsed$FEdate)
```

#----- Generating SCDI Length Variable -----

```
av_pathway_collapsed$FEdate_diag <- as.integer(av_pathway_collapsed$diagdate_new - av_pathway_collapsed$FEdate)
```

```
table(is.na(av_pathway_collapsed$FEdate_diag))  
length(unique(av_pathway_collapsed$FEdate_diag))  
class(av_pathway_collapsed$FEdate_diag)  
summary(av_pathway_collapsed$FEdate_diag)
```

#----- Generating First Event Source Variable -----

```
av_pathway_collapsed$CWTréf_flag <- ifelse(av_pathway_collapsed$CWT_ref_date==av_pathway_collapsed$FEdate,1,0)  
av_pathway_collapsed$CWTfs_flag <- ifelse(av_pathway_collapsed$CWT_FS_date==av_pathway_collapsed$FEdate,1,0)  
av_pathway_collapsed$RTDstart_flag <- ifelse(av_pathway_collapsed$RtD_start_HES_date==av_pathway_collapsed$FEdate,1,0)  
av_pathway_collapsed$RTDref_flag <- ifelse(av_pathway_collapsed$RtD_ref_HES_date==av_pathway_collapsed$FEdate,1,0)  
av_pathway_collapsed$DID6m_flag <- ifelse(av_pathway_collapsed$DID6m==av_pathway_collapsed$FEdate,1,0)  
av_pathway_collapsed$ENDOdate6m_flag <- ifelse(av_pathway_collapsed$HES_ENDOdate6m==av_pathway_collapsed$FEdate,1,0)
```

```
av_pathway_collapsed$CWTréf_flag[is.na(av_pathway_collapsed$CWTréf_flag)] <- 0
```

CAS-SOP: Defining the Secondary Care Diagnostic Interval

```
av_pathway_collapsed$CWTfs_flag[is.na(av_pathway_collapsed$CWTfs_flag)] <- 0
av_pathway_collapsed$RTDstart_flag[is.na(av_pathway_collapsed$RTDstart_flag)] <- 0
av_pathway_collapsed$RTDref_flag[is.na(av_pathway_collapsed$RTDref_flag)] <- 0
av_pathway_collapsed$DID6m_flag[is.na(av_pathway_collapsed$DID6m_flag)] <- 0
av_pathway_collapsed$ENDOfdate6m_flag[is.na(av_pathway_collapsed$ENDOfdate6m_flag)] <- 0
```

```
av_pathway_collapsed$FESource_coded <- (paste(av_pathway_collapsed$CWTref_flag, av_pathway_collapsed$CWTfs_flag, av_pathway_collapsed$RTDstart_flag,
av_pathway_collapsed$RTDref_flag, av_pathway_collapsed$DID6m_flag, av_pathway_collapsed$ENDOfdate6m_flag, sep = ""))
table(av_pathway_collapsed$FESource_coded)
```

```
av_pathway_collapsed <- mutate(av_pathway_collapsed,
  FESource = recode(FESource_coded,
    "000000" = 999,
    "100000" = 1,
    "010000" = 2,
    "001000" = 3,
    "000100" = 4,
    "000010" = 5,
    "000001" = 6,

    "110000" = 7,
    "101000" = 8,
    "100100" = 9,
    "100010" = 10,
    "100001" = 11,

    "011000" = 12,
    "010100" = 13,
    "010010" = 14,
    "010001" = 15,

    "001100" = 16,
    "001010" = 17,
    "001001" = 18,

    "000110" = 19,
    "000101" = 20,

    "000011" = 21,

    "111000" = 22,
    "110100" = 23,
    "110010" = 24,
```

CAS-SOP: Defining the Secondary Care Diagnostic Interval

"110001" = 25,
"101100" = 26,
"101010" = 27,
"101001" = 28,
"100110" = 29,
"100101" = 30,
"100011" = 31,

"011100" = 32,
"011010" = 33,
"011001" = 34,
"010110" = 35,
"010101" = 36,
"010011" = 37,

"001110" = 38,
"001101" = 39,
"001011" = 40,
"000111" = 41,

"111100" = 42,
"111010" = 43,
"111001" = 44,
"110110" = 45,
"110101" = 46,
"110011" = 47,

"101110" = 48,
"101101" = 49,
"101011" = 50,
"100111" = 51,

"011110" = 52,
"011101" = 53,
"011011" = 54,
"010111" = 55,

"001111" = 56,

"111110" = 57,
"111101" = 58,
"111011" = 59,
"110111" = 60,
"101111" = 61,

CAS-SOP: Defining the Secondary Care Diagnostic Interval

```
"011111" = 62,
```

```
"111111" = 63))
```

```
table(av_pathway_collapsed$FESource)
```

```
av_pathway_collapsed <- mutate(av_pathway_collapsed,  
  FESource_grouped = recode(FESource,  
    "1" = "CWT ref",  
    "2" = "CWT FS",  
    "3" = "RtD start",  
    "4" = "RtD ref",  
    "5" = "DID",  
    "6" = "HES",  
    "7" = "other",  
    "8" = "other",  
    "9" = "CWT ref & RtD ref",  
    "10" = "other",  
    "11" = "other",  
    "12" = "other",  
    "13" = "other",  
    "14" = "other",  
    "15" = "other",  
    "16" = "other",  
    "17" = "RtD start & DID",  
    "18" = "other",  
    "19" = "other",  
    "20" = "other",  
    "21" = "other",  
    "22" = "other",  
    "23" = "other",  
    "24" = "other",  
    "25" = "other",  
    "26" = "other",  
    "27" = "other",  
    "28" = "other",  
    "29" = "other",  
    "30" = "other",  
    "31" = "other",  
    "32" = "other",  
    "33" = "other",  
    "34" = "other",  
    "35" = "other",  
    "36" = "other",
```

CAS-SOP: Defining the Secondary Care Diagnostic Interval

```
"37" = "other",  
"38" = "other",  
"39" = "other",  
"40" = "other",  
"41" = "other",  
"42" = "other",  
"43" = "other",  
"44" = "other",  
"45" = "other",  
"46" = "other",  
"47" = "other",  
"48" = "other",  
"49" = "other",  
"50" = "other",  
"51" = "other",  
"52" = "other",  
"53" = "other",  
"54" = "other",  
"55" = "other",  
"56" = "other",  
"57" = "other",  
"58" = "other",  
"59" = "other",  
"60" = "other",  
"61" = "other",  
"62" = "other",  
"63" = "other",  
"999" = "DP missing"))
```

```
table(av_pathway_collapsed$FEsource_grouped)
```

```
#----- Dropping Obs Based on Exclusion Criteria -----
```

```
av_pathway_collapsed <- filter(av_pathway_collapsed, dco != "Y", final_route_coded != "Unknown")
```