



Public Health  
England

Protecting and improving the nation's health

# **National Cancer Registration and Analysis Service**

## **Routes to diagnosis 2006 to 2016**

Technical document

# 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-class 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 are a distinct delivery organisation with operational autonomy to advise and support government, local authorities and the NHS in a professionally independent manner.

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

Prepared by: National Cancer Registration and Analysis Service (NCRAS).

© 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.nationalarchives.gov.uk/ogp/) or email [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk). Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

Published June 2019  
PHE publications  
gateway number: GW-466

PHE supports the UN  
Sustainable Development Goals



# Contents

About Public Health England	2
1.0 Introduction	4
1.1 Overview of the Routes to Diagnosis project	4
2.0 Methods	6
2.1 Overview of the Routes to Diagnosis algorithm	6
2.2 Data sources	6
2.3 De-duplication	9
2.4 Matching algorithms	10
2.5 The Routes to Diagnosis Algorithm	10
2.6 Assigning the Route end-point	10
2.7 Assigning the pathway group code	12
2.8 Assigning the Route start-point	12
2.9 Assigning the detailed Route to Diagnosis code	12
2.10 Assigning the broad Route to Diagnosis category	12
3.0 Analytical techniques	14
3.1 Tumour Grouping	14
3.2 Confidence intervals	14
3.3 Survival analysis	14
3.4 Colour shading	15
4.0 Data quality issues and limitations	16
4.1 Data quality in the fourth iteration	16
4.2 Screening data	16
4.3 Death Certificate Only	17
4.4 Ethnicity	18
5.0 Further methodological development	19
5.1 Outstanding issues within the Routes to Diagnosis algorithm	19
5.2 Expansion of data sources of algorithm	19
Authorship and Acknowledgements	22
Glossary	23
Appendix 1: Routes to diagnosis codes	24
Appendix 3: Start-point codes	28
Appendix 4: Tumour Group National	30

# Introduction

This document summarises the data sources and methodology used for the seventh iteration of the 'Routes to Diagnosis' project covering tumours diagnosed in 2006 to 2016. Results are available on the NCRAS website<sup>1</sup>. The methodology and further study has been published in the British Journal of Cancer<sup>2, 3</sup>.

## Overview of the Routes to Diagnosis project

### Project goals

The questions examined in the 'Routes to Diagnosis' project are described below:

- is it feasible to use routinely available data sources to define the Routes to Diagnosis for patients diagnosed with cancer (for example, whether they present through inpatients, outpatients, screening or via an emergency presentation)?
- if the first is feasible, can the influence of age, sex, ethnicity, deprivation, cancer stage and geographical area of residence on referral routes and pathways be examined?
- is there an association between Routes to Diagnosis and survival for cancer patients?

This document and associated publications<sup>1, 2, 3</sup> demonstrate a positive answer to these 3 questions.

### Technical overview

Administrative Hospital Episode Statistics data are combined with Cancer Waiting Times data, data from the cancer screening programmes and cancer registration data. Using these datasets, every case of cancer registered in England which was diagnosed in 2006 to 2016 is categorised into 1 of 8 Routes to Diagnosis.

### Policy context

The Routes to Diagnosis project supports early diagnosis initiatives whose aim is to promote earlier diagnosis of cancer and thereby improve survival rates and reduce cancer mortality. Successful implementation of such initiatives will make a major contribution to the Independent Cancer Taskforce's goal of achieving world class cancer outcomes in this country.

The Routes to Diagnosis project was the first to explore the feasibility of using routine data to evaluate how cancer patients access the health service for diagnosis and

whether the routes are associated with survival differences. This in turn can be used to inform strategy in terms of improved patient education regarding signs and symptoms, medical practitioner education, and routes of referral. The outputs help to inform awareness and early diagnosis initiatives locally and nationally, ideally resulting in more appropriate referrals and earlier diagnosis of cancer as well as eventually improving the cost-effectiveness of NHS.

## History

The first iteration of the national study was conducted in the summer of 2010 and covered patients diagnosed in 2007<sup>3</sup>.

The second iteration refined the algorithm used and widens the period of data analysed to cancers diagnosed in 2006-2008 (inclusive).

The third iteration incorporates data from 2006 to 2010, covering a broader range of cancer sites.

The fourth iteration covers data from 2006 to 2013 and uses data from the new cancer registration and analysis systems.

The fifth iteration represents a major change in how the algorithm was run and how data were accessed. For the first time the whole process was run within the Cancer Analysis System (CAS) environment with refinements made in defining cohorts and in the speed of processing. This iteration covered 2006-2014 and introduces stage breakdowns to the standard output, as well as using net survival calculations for the first time to line up with the methodology used for official survival statistics.

The sixth iteration covered a 10-year period, 2006 to 2015, with more than 3 million cancer diagnoses. The project builds upon the previous iteration with further refinements aimed at reducing the time between iterations and the run time for the process itself.

The seventh iteration, 2006 to 2016, maintained the data available in previous releases but began a transition with regards to the form the outputs took, with a move to make information easier to access and understand with results being displayed in an online, interactive environment.

The effect of data quality in the seventh iteration is discussed in section 4.1.

## Methods

This section describes the process by which the Routes to Diagnosis algorithm assigns a route to each cancer recorded in the CAS December 2017 frozen snapshot (1712/AV2016), following application of a standard operating procedure to identify cancers.

### Overview of the Routes to Diagnosis algorithm

The algorithm takes as a starting point the date of cancer diagnosis, as defined by the UK and Ireland Association of Cancer Registries (UKIACR) using European Network of Cancer Registries (ENCR) rules<sup>5</sup>. Routine data immediately prior to this date are examined and a series of rules is used to classify the Route to Diagnosis for each case. The routes are categorised in detail by three variables: the end-point, the pathway group, and the start-point. These detailed routes have been aggregated into eight broader categories to facilitate analysis.

It is important to note that patient records being used to describe the Route to Diagnosis may not have a cancer code assigned to them, as the episodes and attendances will have taken place before a cancer diagnosis has been coded. It is therefore not possible to be absolutely certain that the episodes and attendances related to the patient prior to diagnosis were directly related to the process of diagnosis of cancer. However, the frequency of hospital attendance and admission in the period immediately before diagnosis greatly exceeds the 'background' rate, making the assumption that they are related to the cancer diagnosis reasonable<sup>2</sup>.

### Data sources

#### Cancer registration

The CAS holds cancer registration data for the whole of England. The database contains around 13 million cancer registry records. Further information about registration data is available from the NCRAS websites: [www.ncin.org.uk](http://www.ncin.org.uk) and <https://www.ndrs.nhs.uk/>.

All cancer registrations across England between 2006 and 2016 inclusive, with ICD-10 diagnosis codes C00–C97, D00-D09, D13, D15, D27, D29, D33, D33, D35 and D37-D48 (all neoplasms) were obtained from the CAS.

A subset of this data for tumours with ICD-10 diagnosis codes C00-C97 excluding C44 which were diagnosed in calendar years 2006 to 2016 was used for reporting. Other records were excluded from the reporting dataset based on experience in the

previous iteration of Routes to Diagnosis and developments in standard reporting procedures:

- all D codes – with the exception of D05, D06, D090, D32, D330-D332, D333, D334, D352-D354, D42, D430-D432, D433-D434, D437-D439 and D443-D445 – were excluded from the reporting dataset
- the records for patients with non-melanoma skin cancer – as most of these are diagnosed and treated immediately in outpatients or in primary care and Basal Cell Carcinomas – are not subject to the Two-Week Wait (TWW) referral process
- records where the sex code was not female or male – other sex codes are excluded to avoid disclosure as numbers are small
- records where the sex code and cancer site code do not agree – such as male patients with female genital cancer
- records with invalid ages
- records outside of English geographies (defined by country code)
- records with invalid ICD10 site codes
- records not flagged as final registrations
- duplicate records identified by the standard operating procedure

Routes were derived for all tumours fitting the criteria specified, including second and subsequent tumours in the same person (unlike the first iteration of Routes to Diagnosis). Sensitivity analysis conducted for the second iteration showed a small impact on the total if these multiple tumours were excluded: the overall proportion of Emergency Routes would have increased by less than 0.1% and the overall proportion of Unknown Routes would have increased by 0.2%. Other route proportions changed by less than 0.5%. The maximum change in all combinations of route and cancer type on including multiple tumours was 1.7%. with a mean absolute change of 0.2%.

## Hospital Episode Statistics (HES)

Hospital Episode Statistics (HES) is a collection of data sets containing details of all admissions (day case and inpatient) to NHS hospitals in England. It includes details of private patients treated in NHS hospitals; patients that were resident outside England; and of care delivered by treatment centres (including those in the independent sector) funded by the NHS. HES also contain details of all NHS outpatient appointments (attendances for patients who are not formally admitted) in England. It contains admitted patient care data from 1989 onwards, with more than 12 million new records added each year, and outpatient attendance data from 2003 onwards, with more than 40 million new records added each year. Further information about HES is available from the HES online website: <http://content.digital.nhs.uk/hes>.

## Admitted Patient Care (Inpatient and day case) Hospital Episode Statistics

For the national analysis, admitted patient HES for 2003 to 2004 and 2016 to 2017 were used to identify patients with a hospital admission for any cause during this time period. These are commonly referred to as inpatient (IP) HES.

## Outpatient Hospital Episode Statistics

For the national analysis, outpatient (OP) HES for 2003 to 2004 and 2015 to 2017 were used.

## National Cancer Waiting Times

For the national analysis, National Cancer Waiting Times (NCWT) data for 1 July 2004 to 2017 were used. The NCWT system is hosted nationally on NHS.Net (Open Exeter) and allows NHS providers to record data derived from patient care activity. These data are used to monitor performance against the NCWT standards specified in the NHS Cancer Plan 2000 and the Cancer Reform Strategy 2007. As a patient moves through the stages of their treatment pathway, data on referrals, treatments and diagnosis are derived from care records locally. NHS providers are mandated by Data Set Change Notice (DCSN) 20/2008 to collect data concerning all patients covered by the NCWT standards, including patients referred with suspected cancer and patients diagnosed with and treated for new and subsequent cancer. Further information about the NCWT system is available from NHS England's website: [www.england.nhs.uk/statistics/statistical-work-areas/cancer-waiting-times](http://www.england.nhs.uk/statistics/statistical-work-areas/cancer-waiting-times)

## Breast screening

Data from Public Health England (PHE)'s Screening Histories Information Manager (SHIM) system were used to populate the screening data in the cancer registry database, with data being fed in through the Screening QA Reference Centres. This records information for women who had attended breast screening which led to a diagnosis in calendar years 2006 to 2016.

The SHIM data reports screen detected breast cancer cases a few percentage points lower than the previous iterations. This is due to a backlog of cases currently being updated, and differences in the methodology for identifying screen-detected tumours. While the difference for breast cancers is small, in situ screen-detected breast cancers show larger variation. These differences will be fed back to the SHIM system to improve data completeness in future.

## Cervical screening

The screening flag within the CAS was used to determine the screening status of women with cervical cancer. This screening data is collected by cancer registries based on the records held internally as a result of local data exchanges between the cancer registries and the screening Quality Assurance Reference Centres (QARCs). It is known that this underreports the level of screen-detected cervical tumours and variation is predominantly due to variation in collection by the former regional cancer registry offices. The quality of these data has improved in recent years following the creation of a national registration system in 2013.

## Colorectal screening

An offload was provided by NHS Digital, based on data received from the NHS Bowel Cancer Screening Programme. This identified cases of colorectal cancer within the CAS between 2006 and 2016 (inclusive) that had any interaction with the screening programme. Screen-detected cases were identified when a cancer was detected during an open screening episode by the screening programme.

## De-duplication

The CAS dataset was de-duplicated via a standard operating procedure that was applied to cases for England, 2006-2011, by linking to ONS registration data. Data for 2012-2016 were deduplicated by NCRAS. For all years only finalised registrations and cancers with valid ICD10 codes were included.

# Matching algorithms

## Matching algorithm for cancer registration data and HES data

The CAS contains registrations linked to both inpatient and outpatient data. These datasets are linked using an adapted version of NHS digital's Data Linkage and Extract Service linking algorithm.

## Matching cancer registration data and NCWT Data

Records are held in CAS. Records with a referral priority of TWW and a valid Decision to Treat Date were then matched to the records in the registry record in CAS using NHS Number and having a Cancer Diagnosis date between 62 days before and 31 days after the Decision to Treat Date.

## The Routes to Diagnosis Algorithm

The Routes to Diagnosis Algorithm assigns a 3-part code to each tumour based on the inpatient and outpatient HES data, as described below in sections 2.6-2.9. This 3-part code is either mapped to 1 of 7 broader route categories or the presence of Screening or Cancer Waiting Times data can take precedence and cause the final route to be a TWW or screen-detected route, as described in section 2.10.

## Assigning the route end-point

A specific inpatient or outpatient episode was identified in HES as the 'end-point' of the route by its proximity to the date of diagnosis. The end-point was assumed to be the clinical care event that led most immediately to diagnosis. Where both inpatient and outpatient activity occurred on the date of diagnosis the inpatient episode was defined as the end-point of the route. Otherwise, if there was an episode within 28 days prior to the date of diagnosis then this was assigned as the end-point of the route, with inpatient episodes taking precedence over outpatient episodes and the most recent episode taking precedence if there were multiple episodes. If there was no HES activity within 28 days of diagnosis then the most recent episode within 6 months (inpatient or outpatient) was used as the end-point of the route.

The following end-point codes were assigned:

### Special cases (SC)

Patients with a cancer diagnosis date on the same day as an inpatient admission date and an outpatient attendance date, or whose closest HES episodes to diagnosis are an inpatient and outpatient record occurring on the same date. These are a special case of inpatient diagnosis.

### Inpatient diagnosis (IP)

Patients with a cancer diagnosis date related to a preceding inpatient HES episode (excluding patients already defined as special cases). An IP is defined where the cancer diagnosis date is within the start and end of an episode. In addition, due to the potential for diagnosis to be confirmed following a relevant inpatient episode, a cancer diagnosis date that is within 6 months after the end of an episode and with no outpatient episode between would also be regarded as an IP.

### Outpatient diagnosis (OP)

Patients with no inpatient HES episode preceding the cancer diagnosis date (as defined above) but with an outpatient HES attendance preceding the cancer diagnosis date or with an IP elective admission, or were emergencies via an outpatient clinic, or were unresolved inpatient transfers.

### Unknown (UN)

Unable to match cancer diagnosis date to any informative inpatient or outpatient HES episode within the valid timeframe. It is likely that, for these patients, the cancer diagnosis date was obtained from pathology records only, indicating diagnosis or treatment that only took place outside of a hospital setting (such as NHS patients seen in primary care, independent treatment centres or a community setting, and private patients seen and treated only in private hospitals).

### Death Certificate Only diagnosis (DCO)

The cancer registry receives a small number of cancer-related death notifications, for which, despite extensive enquiries, they are unable to obtain additional information to register the disease details fully. This registration is regarded as Death Certificate Only (DCO) and the date of diagnosis is the same as that of the date of death.

## Assigning the pathway group code

Each tumour was assigned a pathway group code based on the presence of inpatient and outpatient HES data as detailed in Table 2.1.

**Table 2.1: Pathway Group codes**

Pathway group	Description
A	Inpatient HES only within 6 months prior to diagnosis
B	Outpatient HES only within 6 months prior to diagnosis
C	Special case, an inpatient elective or Emergency via outpatient clinic, and there is outpatient HES within 6 months prior to diagnosis
D	There are no HES data 6 months prior to diagnosis
E	No HES data at all prior to diagnosis

## Assigning the route start-point

The start-point is determined by working backwards from the end-point as shown in Appendix 2. The characteristics of this start-point lead to a categorisation of route:

- routes that originated in an outpatient attendance use the outpatient source of referral of that attendance as the start-point code
- routes that originated in an inpatient episode use the inpatient method of admission as the start-point code
- routes where inpatient or outpatient data were unavailable the start-point codes may be assigned as null or unknown (this also includes DCOs)

A list of all possible start-point codes is provided in Appendix 3.

## Assigning the detailed Route to Diagnosis code

For each patient, a route end-point, the pathway group and the route start-point were derived and an overall detailed route code was defined by the concatenation of these 3 codes in the specific order: end-point–pathway group-start-point (for example, IP-02-O03). This resulted in a total of 70 distinct routes to diagnosis codes, listed in Appendix 1.

## Assigning the broad route to Diagnosis category

To be useful for analytical purposes these must be aggregated into a manageable number of broader categories. Upon examination 2 categories were identified which represent qualitatively different routes – screen-detected and DCO. Three routes reflect the urgency of referral (Emergency, TWW referral and other GP referral). Two further routes represent cases for which the route apparently started in secondary

care (Inpatient Electives and Other Outpatients) and, finally, one reflects cases with no useful information available on the Route to Diagnosis (UNs). These 8 groups are detailed below:

- GP referral – includes routine and urgent referrals where the patient was not referred under the TWW referral route
- Two-Week Wait – urgent GP referrals with a suspicion of cancer
- emergency presentation – an emergency route via accident and emergency (A&E), emergency GP referral, emergency transfer, emergency admission or attendance
- other outpatient – an elective route starting with an outpatient appointment that is either a self-referral, consultant to consultant referral, other or unknown referral (these referrals would not include patients originally referred under the TWW referral route)
- screen-detected – flagged by the cancer registry as detected via the breast or cervical screening programmes
- inpatient elective: where no earlier information can be found prior to admission from a waiting list, booked or planned
- DCO: diagnosis by death certificate only
- Unknown: no relevant data available from IP or OP HES or from NCWT or screening

The table in Appendix 1 was used to allocate route categories from HES data.

After routes were allocated to each case from the HES data the screening and CWT data were examined. Where a case could be linked to a CWT urgent referral for suspected cancer it was categorised as a TWW route, unless the route categorised using the HES data was an Emergency Presentation with an admission date within 28 days prior to the Decision to Treat date. Where the case could be linked to a screening event the route was categorised as Screening. If both were possible then a Screen Detected route took priority over a TWW route.

A case was linked to a CWT referral where a TWW had a Decision to Treat date within 62 days prior to or 31 days after the date of diagnosis. A case was linked to a breast screening event where the woman was identified by the SHIM data to be a screen detected case. A case was identified as screen detected by the National Bowel Screening Programme where a case was linked to an event relating to a screening episode with an identification of cancer detected by the programme. For cervical screening data the determination that the case was screen detected had been made by the cancer registry and no matching by date was performed.

## Analytical techniques

This section details analytical methods used to interpret the outputs from the Routes to Diagnosis algorithm on the CAS.

### Tumour grouping

For the analysis, 56 tumour types or groups were identified, primarily based on advice from site-specific expert advisory groups and suitability for the various breakdowns published. The list of tumour types by International Classification of Disease (ICD-10) codes is provided in Appendices 4.1 and 4.2.

### Confidence intervals

Binomial confidence intervals for proportions of cancers diagnosed via a particular Route to Diagnosis are calculated using the Wilson score method<sup>6</sup>. Confidence intervals for survival analysis were calculated as part of the strs algorithm as defined below.

### Survival analysis

Net survival is the probability of surviving cancer in the absence of other causes of death. Estimates were produced by using life tables wherein mortality of cancer patients was compared to the general population matched by age, sex, socio-economic status and geographic region.

One-year survival estimates were calculated using the stns command developed by Clerc-Urmés, Grzebyk and Hédelin<sup>10</sup> in conjunction with a cohort approach and the Pohar-Perme net survival estimator<sup>11</sup> run within Stata version 14.

Net survival has been calculated on 5 year rolling cohorts, with the exception of 36-month survival (2011-2015).

The following criteria are used to identify the patients that are eligible to be included in the analysis:

- patients should have a unique identifier
- patients should have a complete date of birth and be aged between 15 and 99 at diagnosis
- patients who have died should have a complete registered date of death
- patients should have a complete date of cancer diagnosis
- patients should have a known sex
- patients should have a known date of being recorded as alive or dead

- patients should be residents of England and have a valid postcode for their usual place of residence
- tumours should be malignant and be newly diagnosed in the studied cohort and be a primary tumour
- cancers of the blood (lymphomas, leukaemias and myelomas) should not occur in a solid tumour
- patients are not excluded if they had further primary tumours of the cancer of interest later in the period of interest, nor if they had any primary tumour of another cancer site diagnosed in the period of interest, nor if they had any type of primary tumour diagnosed before or after the period of interest
- patients are excluded if they are registered via Death Certificate Only (DCO) where the first confirmed record of a tumour occurs after the death of the patient
- the sequence of dates should be valid (for example a patient should not be diagnosed before they are born)

As a result of these criteria, cohort numbers for each route will not necessarily total up to the all routes number. For cohorts with very few deaths, estimates are suppressed for longer survival times. For example, the screen detected route for female breast cancer only has estimates for 1-month survival, as very few deaths occur as survival time increases.

## Colour shading

Tabular data is commonly presented with the background colour in each tabular cell related to the magnitude of the proportion in the cell. The extreme values of the background colours are set by the extreme values of the tabular data. Depending on the context of the table the extreme values might be those in the whole table or in one particular row or column. The colouring of each table should be considered a subjective 'guide to the eye' rather than having a fixed relationship to the magnitude of the data.

## Data quality issues and limitations

This section outlines data quality issues in the raw data used.

### Data quality in the fourth iteration

An issue has been identified with the inclusion of cancer registrations diagnosed in areas covered by West Lancashire CCG, and to a lesser extent Eastern Cheshire, South Cheshire, and Vale Royal. This will affect local authority data and regional data that cover this area. Cancer cases may be missing from the cohort of patients diagnosed prior to 2008. This only relates to a relatively small number of cases and so will not impact greatly on the national figures, but may mean that cancer incidence is significantly underestimated in these areas.

An improvement in the completeness of HES data linked to cancer registrations has led to HES data being available for more tumours than in the earlier iterations of RtD. Previously, only Admitted Patient Care (APC) HES data, which includes inpatient and day case activity, were linked to cancer registrations where one of the episodes contained an ICD10 C00-C97, D00-D48 or O01 code in one of the diagnosis fields. With approval from the Confidentiality Advisory Group (CAG, formerly National Information Governance Board NIGB), APC and outpatient HES data are now supplied on the basis of a cancer registration existing for a patient. If an HES record can be linked to a cancer registration, then all HES data are supplied.

As a result, HES data are available for more tumours. Comparisons between this iteration of Routes to Diagnosis and the previous iteration, for tumours diagnosed in 2006-2015, shows that this has led to a stabilisation in the proportion of tumours assigned to the Unknown Route which remains at 3.7%. The same pattern is seen in the proportion of inpatient electives which remain at 2.3% and TWW referrals which stayed at 30.9%.

The methodology regarding the application of HES data has not changed from the previous iteration.

### Screening data

An analysis of completeness of screening flags for England provided by National Disease Registration was undertaken, see Table 4.1. The breakdown by site shows a variation in the percentage of screen-detected records assigned by the registration service and the Routes to Diagnosis algorithm using the data from the screening programmes. For breast and colorectal cancers the project uses the data provided by the respective NHS Cancer Screening Programmes. Cervical screening data are not available directly at present, but it is hoped that this data set will be added in the future.

**Table 4.1: The number of records for England patients against each of the breast and cervical ICD-10 groupings (C50, C53, D05 and D06) by screen detected flag, 2006-2016**

Cancer site	Registry screen-detected flag			Routes screen-detected flag		Count
	Yes	No	Other value	Yes	No	
Cervix	21%	35%	44%	28%	72%	27,825
Cervix (in-situ)	0%	0%	100%	22%	78%	265,629
Colorectal	6%	16%	77%	8%	92%	370,712
Breast	28%	32%	40%	28%	72%	471,125
Breast (in-situ)	51%	15%	34%	53%	47%	61,023

### Death Certificate Only

Patients who were registered as a DCO on the CAS and could not be matched to any of the data sources referenced in Section 2.2 above were assigned a DCO route grouping. However, there were patients registered as DCOs where additional information was found in inpatient and/or outpatient HES data which allowed these patients to be assigned a different route grouping, see Tables 4.3 and 4.4. This finding has important incidental implications for reducing the DCO rate recorded by the cancer registry. All tables below show the number of records as opposed to the number of distinct patients, which includes all records in the analysis (for instance, it does not exclude multiples).

**Table 4.3: Comparison of the number of records assigned to the different routes to diagnosis against the number of records that have been flagged by the cancer registry as being DCO or non-DCO**

Count of records		Registry DCO flag		
Route to Diagnosis		Yes	No	Total
DCO		11,013	0	11,013
Not DCO total	Emergency presentation	27,066	688,582	715,648
	GP referral	9,553	1,176,821	1,186,374
	Inpatient elective	535	88,448	88,983
	Other outpatient	6,208	396,218	402,426
	Screening	11	259,767	259,778
	TWW	208	1,022,642	1,022,850
	Unknown	303	154,740	155,043
Not DCO total		43,884	3,787,218	3,831,102
Total records		54,897	3,787,218	3,842,115

**Table 4.4: Comparison of percentage of records assigned to the DCO and non-DCO Routes to Diagnosis groupings against the percentage of records that have been flagged by the cancer registry as being DCO or non-DCO**

Percentage of records Route to Diagnosis	Registry DCO Flag		
	Yes	No	Total
DCO	0.3%	0.0%	0.3%
Not DCO	1.1%	98.6%	99.7%
Total records	1.4%	98.6%	100.0%

## Ethnicity

Ethnicity data for positive ethnicity recording were gained from the 3 available HES sources (A&E, Admitted Patient Care and outpatient datasets) for the years 2006 to 2011. These data were combined and linked back through to the appropriate data row in the CAS tables. From 2012, ethnicity was taken from the cancer registration fields on CAS. Frequencies of ethnic grouping recording were defined, from which a final most commonly recorded ethnicity was gained. Ethnicity groups included are as follows: Asian, Black, Chinese, Mixed, White, Other ethnic group (other ethnicity) and Unknown, where no ethnicity could be derived.

**Table 4.5: Proportion of recorded ethnicity by ethnic group**

Ethnicity	Asian	Black	Chinese	Mixed	Other	White	Unknown	Total recorded
Proportion of patients	2%	2%	0%	0%	1%	89%	6%	94%

## Further methodological development

This section records points noted in the development of the Routes to Diagnosis algorithm to further improve or develop it.

### Outstanding issues within the Routes to Diagnosis algorithm

Several minor issues were noted during the development of the Routes to Diagnosis algorithm for the second iteration. These issues remain in the current iteration. They are described below with steps that might be taken to resolve them.

#### DNA and cancelled status of outpatient episode

A small percentage of outpatient episodes, while present in the dataset, are coded as DNA (indicating that the patient did not attend) or cancelled. The project team decided not to remove these episodes in the belief that information contained in the episode might still be relevant to the patient's Route to Diagnosis. This will be reviewed as part of further Routes to Diagnosis development work.

#### Multiple outpatient attendances on same day

A small percentage of outpatients attendances occur on the same day as another outpatient appointment. In these cases the temporal order was assigned randomly for purposes of deciding which was closer to the time of diagnosis. As of the fifth iteration, these IDs used for random assignment were updated and changed in the latest version of HES, resulting in half of these assignments switching order. While this results in a very minor change, it does result in some inconsistency with previous iterations.

#### Expansion of data sources of algorithm

The Routes to Diagnosis algorithm relies on Cancer Registration data, plus in- and outpatient HES data, Cancer Waiting Times data, and data from the Breast, Cervical (via the cancer registries) and Colorectal screening services. Including further data sources may add to the robustness or utility of the algorithm. With the full conversion of the algorithm in to the CAS environment this process can be more easily explored. It is hoped to trial proxy data to examine primary care pathways in future.

#### Expand to include Accident and Emergency data

A&E HES data may provide more complete information on Emergency Presentations or enable them to be analysed at a more granular scale. The feasibility of building A&E HES data into the algorithm should be explored.

## Expand to include Diagnostic Imaging Dataset (DID)

Diagnostic imaging carried out in secondary care should be picked up by the Routes to Diagnosis algorithm as part of an outpatient attendance or inpatient episode. However, imaging conducted in primary care will not currently be captured. DID data are in the early stages of use in analysis and quality testing and their incorporation will be explored once the datasets accuracy has been assured.

## Expand to include Primary Care data

The secondary care setting is the focus of the datasets currently used by the algorithm (except screening). Adding primary care data would allow the parts of the Route to Diagnosis which takes place in Primary Care to be mapped. While there are not presently any Primary Care datasets which have complete national coverage the feasibility of including primary care data, such as that collected as part of the National Cancer Diagnosis Audit (NCDA), in the algorithm should be explored.

## References

1. Routes to Diagnosis 2006-2016 available online at [www.ncin.org.uk/](http://www.ncin.org.uk/)
2. Routes to Diagnosis for cancer - Determining the patient journey using multiple routine datasets (2012). L Elliss-Brookes et al. British Journal of Cancer Vol 107, No 8.
3. Emergency presentation of cancer and short-term mortality (2013). S McPhail et al. British Journal of Cancer. doi: 10.1038/bjc.2013.569. [Epub ahead of print]
4. NCIN Data Briefing, Routes to Diagnosis available online at [http://www.ncin.org.uk/publications/data\\_briefings/routes\\_to\\_diagnosis.aspx](http://www.ncin.org.uk/publications/data_briefings/routes_to_diagnosis.aspx)
5. Tyczynski J, Démaret E, Parkin D, eds. (2003) Standards and guidelines for cancer registration in Europe. IARC Technical Publication. 40: 69–73.
6. APHO Technical Briefing 3: Commonly used public health statistics and their confidence intervals. Available online at: [www.apho.org.uk/resource/view.aspx?RID=48457](http://www.apho.org.uk/resource/view.aspx?RID=48457)
7. Analytical Tools for Public Health: Funnel plot for proportions and percentages <http://www.apho.org.uk/resource/view.aspx?RID=47241>
8. [http://www.stata.com/meeting/uk09/uk09\\_gini\\_forni.pdf](http://www.stata.com/meeting/uk09/uk09_gini_forni.pdf)
9. Funnel plots for comparing institutional performance, Spiegelhalter, 2005: [http://www.medicine.cf.ac.uk/media/filer\\_public/2010/10/11/journal\\_club\\_-\\_spiegelhalter\\_stats\\_in\\_med\\_funnel\\_plots.pdf](http://www.medicine.cf.ac.uk/media/filer_public/2010/10/11/journal_club_-_spiegelhalter_stats_in_med_funnel_plots.pdf)
10. Net survival estimation with stns (2014). I Clerc-Urmès, M Grzebyk, G Hédelin. The Stata Journal 14. Number 1: 87–102 <http://www.stata-journal.com/article.html?article=st0326>
11. On Estimation in Relative Survival (2011). M Pohar Perme, J Stare, J Estève, Journal of the International Biomtrics Society. Volume 68, Issue 1, March 2012 , Pages 113–120. <http://onlinelibrary.wiley.com/doi/10.1111/j.1541-0420.2011.01640.x/abstract>

## Authorship and acknowledgements

The Routes to Diagnosis seventh iteration project team consisted of Tom Bacon, John Broggio, Sam Winters and Sean McPhail with support from Charlotte Eversfield and Jen Than.

The Routes to Diagnosis sixth iteration project team consisted of Tom Bacon, John Broggio, Sam Johnson and Sean McPhail with support from Lucy Irvine and Bukky Juwa.

The Routes to Diagnosis fifth iteration project team consisted of John Broggio, Sam Johnson, Sean McPhail and Miguel Torres.

The Routes to Diagnosis fourth iteration project team consisted of Isabella Carneiro, Matt Greenslade, Luke Hounsome, Sam Johnson and Sean McPhail.

The Routes to Diagnosis third iteration project team consisted of Alex Ives, Matt Greenslade, Sam Johnson, Sean McPhail, and Jon Shelton.

The Routes to Diagnosis second iteration project team consisted of Lucy Elliss-Brookes, Alex Ives, Matt Greenslade, Sean McPhail, Mike Richards, and Jon Shelton.

The project team is grateful for the help and support of Chris Carrigan, the late Brian Cottier, Shan Cheung, Paul Eves, Sara Hiom, Olive Kearins, Tariq Malik, Jodie Moffat, Eva Morris, Andy Pring, Andy Smith, James Thomas, Catherine Thomson, Julia Verne and the other staff of the National Cancer Registration and Analysis Service, PHE Knowledge and Intelligence teams, Cancer Research UK, and the National Cancer Services Analysis Team.

# Glossary

## Pathway Group

A classification that is created for each tumour according to the presence or absence of inpatient and outpatient HES data in the 6 months prior to diagnosis.

## Route to Diagnosis

A Route to Diagnosis is defined as the sequence of interactions between the patient and the healthcare system which lead to a diagnosis of cancer, based on the end point, the pathway and the referral route into secondary care. Depending on context it might either be a detailed route, for example IP-C-O4, or a broad summary route, such as Emergency Presentation.

## Route start point

The start point is the first recorded clinical care event that the Route to Diagnosis Algorithm picks up.

## Route end point

The end point was assumed to be the clinical care event that led most immediately to diagnosis.

## Appendix 1: Routes to Diagnosis codes

A list of all Routes to Diagnosis codes is provided in Table A1.1. The route code is in the form of route end point – pathway group– start-point.

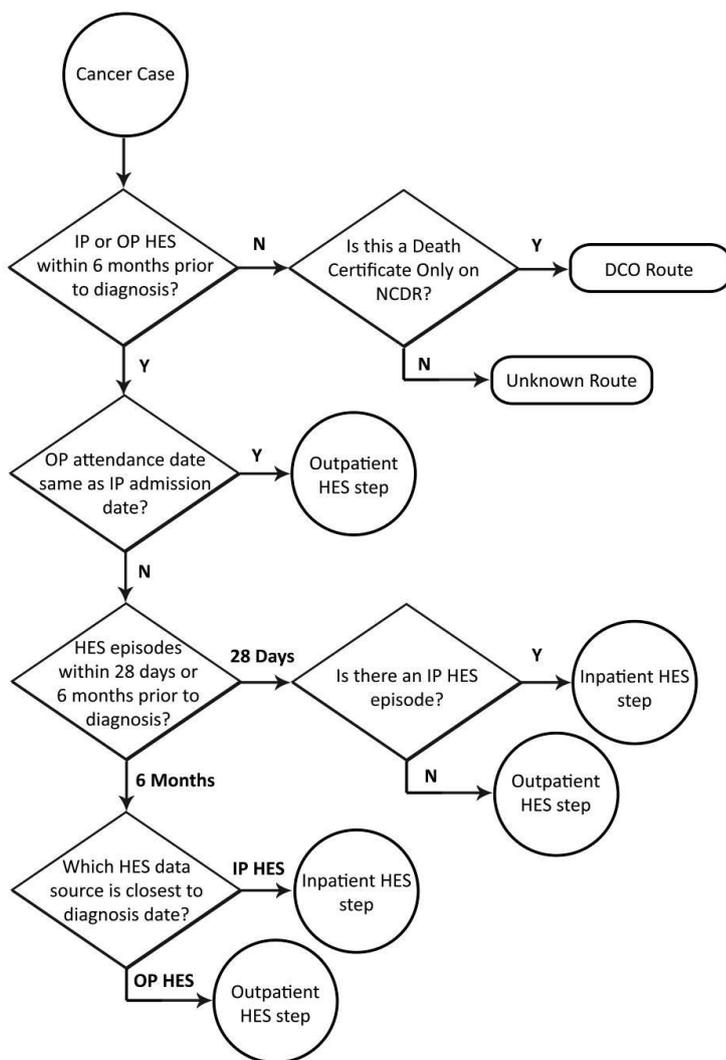
**Table A1.1: Route to Diagnosis codes**

No.	Route code	Route group	No.	Route code	Route group
1	DC-D-DCO	DCO	43	OP-B-O02	Other outpatient
2	DC-E-DCO	DCO	44	OP-B-O03	GP referral
3	IP-A-I11	Inpatient Elective	45	OP-B-O04	Emergency presentation
4	IP-A-I12	Inpatient Elective	46	OP-B-O05	Other outpatient
5	IP-A-I13	Inpatient Elective	47	OP-B-O06	Other outpatient
6	IP-A-I21	Emergency presentation	48	OP-B-O07	Other outpatient
7	IP-A-I22	Emergency presentation	49	OP-B-O08	Other outpatient
8	IP-A-I23	Emergency presentation	50	OP-B-O10	Emergency presentation
9	IP-A-I24	Emergency presentation	51	OP-B-O11	Other outpatient
10	IP-A-I28	Emergency presentation	52	OP-B-O12	GP referral
11	IP-A-I2A	Emergency presentation	53	OP-B-O13	Other outpatient
12	IP-A-I2B	Emergency presentation	54	OP-B-O14	Other outpatient
13	IP-A-I2D	Emergency presentation	55	OP-B-O15	Other outpatient
14	IP-A-I31	Inpatient Elective	56	OP-B-O16	Other outpatient
15	IP-A-I32	Inpatient Elective	57	OP-B-O17	Screening
16	IP-A-I81	Inpatient Elective	58	OP-B-O92	Other outpatient
17	IP-A-I82	Inpatient Elective	59	OP-B-O93	Other outpatient
18	IP-A-I83	Inpatient Elective	60	OP-B-O97	Other outpatient
19	IP-A-I84	Inpatient Elective	61	OP-B-O99	Unknown
20	IP-A-I98	Unknown	62	SC-C-null	Unknown
21	IP-A-I99	Unknown	63	SC-C-O01	Emergency presentation
22	IP-C-O01	Emergency presentation	64	SC-C-O02	Other outpatient
23	IP-C-O02	Other outpatient	65	SC-C-O03	GP referral
24	IP-C-O03	GP referral	66	SC-C-O04	Emergency presentation
25	IP-C-O04	Emergency presentation	67	SC-C-O05	Other outpatient
26	IP-C-O05	Other outpatient	68	SC-C-O06	Other outpatient
27	IP-C-O06	Other outpatient	69	SC-C-O07	Other outpatient
28	IP-C-O07	Other outpatient	70	SC-C-O08	Other outpatient
29	IP-C-O08	Other outpatient	71	SC-C-O10	Emergency presentation
30	IP-C-O10	Emergency presentation	72	SC-C-O11	Other outpatient
31	IP-C-O11	Other outpatient	73	SC-C-O12	GP referral

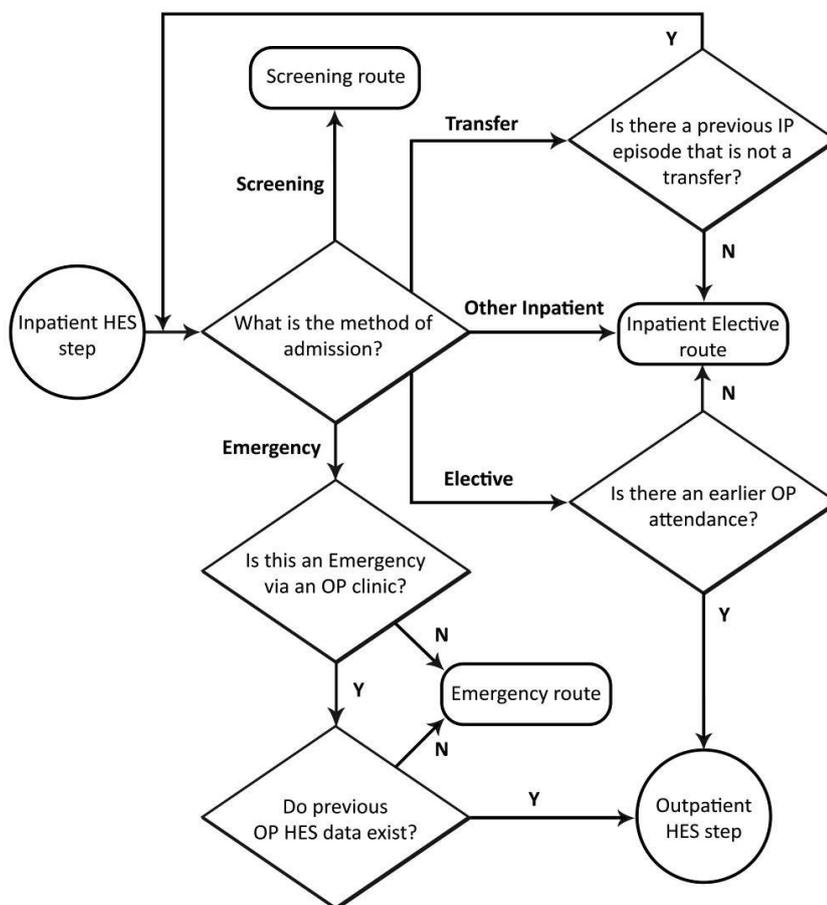
32	IP-C-O12	GP referral	74	SC-C-O13	Other outpatient
33	IP-C-O13	Other outpatient	75	SC-C-O14	Other outpatient
34	IP-C-O14	Other outpatient	76	SC-C-O15	Other outpatient
35	IP-C-O15	Other outpatient	77	SC-C-O16	Other outpatient
36	IP-C-O16	Other outpatient	78	SC-C-O17	Screening
37	IP-C-O17	Screening	79	SC-C-O92	Other outpatient
38	IP-C-O92	Other outpatient	80	SC-C-O93	Other outpatient
39	IP-C-O93	Other outpatient	81	SC-C-O97	Other outpatient
40	IP-C-O97	Other outpatient	82	SC-C-O99	Unknown
41	OP-B-null	Unknown	83	UN-D-UNK	Unknown
42	OP-B-O01	Emergency presentation	84	UN-E-UNK	Unknown

## Appendix 2: Algorithmic flow diagrams

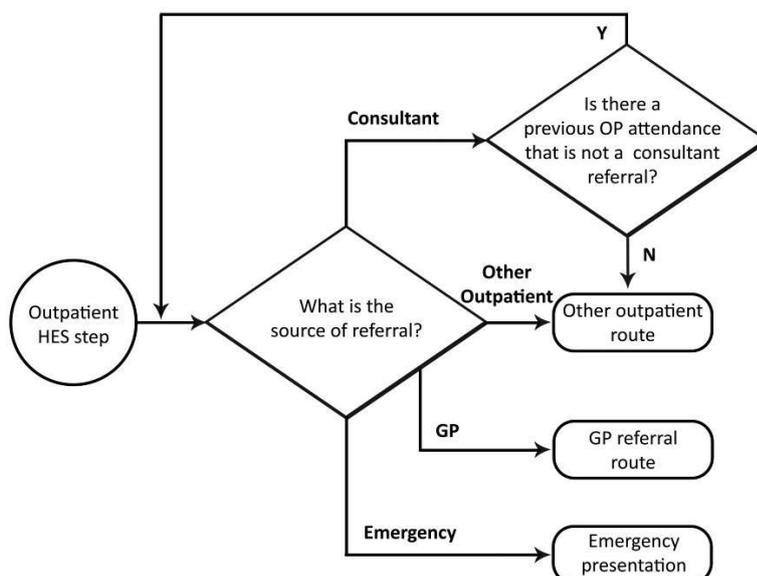
Figure A2.1: Flow diagram for allocating the end point of the route using inpatient and outpatient HES data



**Figure A2.2: Flow diagram for finding the start point or prior step for an inpatient step in a route**



**Figure A2.3: Flow diagram for finding the start point or prior step for an outpatient step in a route**



## Appendix 3: Start-point codes

A list of all 'start-point' codes is provided in Table A3.1. Codes that commence with an 'I' indicates an inpatient method of admission while an 'O' indicates an outpatient source of referral.

**Table A3.1: Start-point codes**

Start point code	Start point Description
DCO	DCO
I11	Elective: from waiting list
I12	Elective: booked
I13	Elective: planned
I21	Emergency: via Accident and Emergency (A&E) services, including the casualty department of the provider
I22	Emergency: via general practitioner (GP)
I23	Emergency: via Bed Bureau, including the Central Bureau
I24	Emergency: via consultant outpatient clinic
I28	Emergency: other means, including patients who arrived via the A&E department of another healthcare provider
I31	Maternity: where the baby was delivered after the mothers admission
I32	Maternity: where the baby was delivered before the mothers admission
I81	Transfer of any admitted patient from another hospital provider other than in an emergency; this does not include admissions to high security psychiatric hospitals (HSPH)
I82	Other: babies born in health care provider
I83	Other: babies born outside the health care provider, except when born at home as intended
I84	Admission by the admission panel of an HSPH; patient not entered on the HSPH admissions waiting list (not valid for admissions after 31 March 2002)
I98	Not applicable (e.g. other maternity event)
I99	Not known
O01	Following an emergency admission
O02	Following a domiciliary visit
O03	Referral from a general medical practitioner
O04	Referral from an accident and emergency department
O05	Referral from a consultant, other than in an accident and emergency department
O06	Self referral
O07	Referral from prosthetist
O08	Other source of referral
O10	Following an accident and emergency attendance
O11	Other

O12	Referral from GP with special interest
O13	Referral from a specialist nurse (secondary care)
O14	Referral from an allied health professional
O15	Referral from an optometrist
O16	Referral from an orthopist
O17	Referral from a national screening programme
O92	General dental practitioner
O93	community dental service
O97	Other - not initiated by the consultant responsible for the consultant outpatient episode
O99	Not known
UNK	Unknown

## Appendix 4: Tumour Group National

**Table A4.1: Tumour categories with associated ICD-10 codes**

Cancer site/group	ICD10 codes included	Available results	
		Overall incidence, overall incidence metrics and survival estimates*	Incidence by year
Anus	C21	♦	
Bladder	C67	♦	♦
Bladder (in-situ)	D090	♦	♦
Brain	C71, D330-D332, D430-D432	♦	♦
Intracranial endocrine	C751-C753, D352-D354, D443-D445	♦	
Meninges	C70, D32, D42	♦	♦
Spinal cord and Cranial nerves	C720-C725, D333, D334, D433-D434	♦	
Other CNS and intracranial tumours	C720-C725, C75, D333-D334, D352-D354, D433-D434, D443-D445		♦
Breast	C50	♦	♦
Breast (in-situ)	D05	♦	♦
Cancer of Unknown Primary	C77, C78, C79, C80	♦	♦
Cervix	C53	♦	♦
Cervix (in-situ)	D06	♦	♦
Colorectal	C18, C19, C20	♦	♦
Gallbladder	C23	♦	
Head and neck - Eye	C69	♦	
Head and neck - Hypopharynx	C12, C13	♦	
Head and neck - Larynx	C32	♦	♦
Head and neck - Nasopharynx	C11	♦	
Head and neck - Oral cavity	C02, C03, C04, C06	♦	♦
Head and neck - Oropharynx	C01, C09, C10	♦	♦

Head and neck - Palate	C05	♦	
Head and neck - Salivary glands	C07, C08	♦	
Head and neck - Thyroid	C73	♦	♦
Head and Neck - non specific	C00, C14, C31	♦	♦
Head and neck - Other (excl. oral cavity, oropharynx, larynx & thyroid)	C05, C07, C08, C11, C12, C13		♦
Heart, Mediastinum and Pleura	C38	♦	
Hodgkin lymphoma	C81	♦	♦
Non-Hodgkin lymphoma	C82, C83, C84, C85	♦	♦
Kidney	C64	♦	♦
Other and unspecified urinary	C65, C66, C68	♦	♦
Leukaemia: acute lymphoblastic	C910	♦	
Leukaemia: acute myeloid	C920, C924, C925 C930, C940, C942	♦	♦
Leukaemia: chronic lymphocytic	C911	♦	♦
Leukaemia: chronic myeloid	C921	♦	
Leukaemia: other (all excluding AML and CLL)	C910, C921		♦
Other haematological malignancies	C88, C912-C919, C922, C923, C927-C929, C931-C939, C943-C947, C95, C96	♦	
Other haematological malignancies	C88, C912-C919, C922, C923, C927-C929, C931-C939, C943-C947, C95, C96		♦
Liver (excl intrahepatic bile duct)	C220, C222-C229	♦	
Biliary tract cancer	C221, C240, C248-C249	♦	
Liver	C221-C229		♦
Lung	C33, C34	♦	♦
Melanoma	C43	♦	♦
Mesothelioma	C45	♦	♦
Multiple myeloma	C90	♦	♦
Nasal Cavity and Middle Ear	C30	♦	

Oesophagus	C15	♦	♦
Other malignant neoplasms	C241, C26, C37, C39, C46, C47, C58, C63, C74, C750, C754-C759, C728-C729, C76, C97	♦	
Other malignant neoplasms	C17, C21, C23, C26, C30, C37, C38, C39, C46, C47, C52, C58, C60, C63, C69, C728-C729, C74, C75, C76, C97		♦
Ovary	C56, C57	♦	♦
Pancreas	C25	♦	♦
Penis	C60	♦	
Prostate	C61	♦	♦
Sarcoma: Bone	C40, C41	♦	♦
Sarcoma: connective and soft tissue	C48, C49	♦	♦
Small Intestine	C17	♦	
Stomach	C16	♦	♦
Testis	C62	♦	♦
Uterus	C54, C55	♦	♦
Vagina	C52	♦	
Vulva	C51	♦	♦

\* Note for survival estimates some sites have been relabelled to align with the National Statistics. These changes are as follows:

Brain	Brain: invasive and benign
Multiple myeloma	Myeloma
Ovary	Ovary: inc. NOS gynae