

Get Data Out technical documentation

Routes to Diagnosis

GDO 0022, 2021-05-04

Background

Routes to Diagnosis data categorise the pathway a patient took to their diagnosis into one of eight Routes, assigned by a published methodological process and run routinely within CAS. The output is a single table with the Route and Route code that can be linked through tumourid.

1. Datasets used

The end of year AV20XX dataset in CASREF01 are used to produce these figures. Routes data used two standard datasets and one lookup table:

- AV2018.AT_Tumour_England – This was used to select the cohort, tumour and demographic details as well as provide links to other datasets via the ID fields
- AV2017.RTD2017 – This was used to select the Route to Diagnosis, linked to the cohort via tumourid. AV2017 was used in this instance as 2018 data were not available. This dataset was created following the [Routes to Diagnosis 2006 to 2016 technical document](#).

The formal publication explaining the Routes methodology: [Routes to Diagnosis for cancer – Ellis-Brookes et al.pdf](#)

2. Cohort

The cohort was defined to align with the cohort specified in [GDO_0021_tech_doc_Incidence](#), which in turn is generated from the methodology and standard restrictions in the [CAS SOP #1 – Counting Cancer Cases](#). This SOP was followed, with the additional exclusions of testicular tumours in female patients and ovarian tumours in male patients for tumours whose ICD code starts with D. (The SOP handles C-coded tumours but does not yet perform any exclusions on D-coded tumours: sex-exclusions are performed on all tumours for our dataset). The cohort is created from AV2018.at_tumour_england. Cases not meeting the SOP criteria are excluded

from the cohort. Once the restrictions have been applied the cohort is refined using diagnosis year and site codes to select the relevant cohort of interest.

3. Variation from published figures

The figures published in the GDO project will vary slightly from the official Routes to Diagnosis figures published by NCRAS. This is due to a difference in the cohort being used to create the denominator. For published Routes work this cohort only includes cases that were run through the Routes to Diagnosis algorithm and have been classified into a Route. This algorithm has some exclusion criteria around data quality for all datasets that may drop some cases but will also be subject to registry creep – with cases registered after the algorithm has been run not receiving a Route. The RTD2017 table was run on a cohort built from AV2017 for 2017. The 2013-2017 cohort from AV2018 includes some cases registered late that were not included in the RTD2017 table. These cases are marked as “Not classified”, signifying that the Routes algorithm has not touched them. They are not assigned to the Unknown Route – this only happens for tumours that have been run through the algorithm. The addition of this classification was chosen so the cohort figures used here match with other GDO metrics.

The sites included here also vary from published Routes to Diagnosis data – so variation in results may also be down to the inclusion of different ICD10 codes, despite site names being identical in some cases. Age and geographic cuts may also differ.

4. Other caveats

Where no standard screening programme for the cancer sites in the initial data release exists, the screen detected Route is removed. Some but not all sarcomas are eligible for screening, but some are screen detected and are counted as such.

Routes for 2018 diagnoses are not included here, as Routes to Diagnosis does not yet cover this year.

Where a group size is very small, data are not available as a measure to protect patient confidentiality. This is indicated with ‘.k’.

5. QA

The QA was done internally by the Get Data Out team.

6. Alterations to technical specification

Column names for the Routes data have been changed since our previous publication. We know this may make difficulties for people loading the data into standard templates, and we aim to minimise this. The rational for these changes are:

- Addition of 'percentage' into each Route lower and upper confidence interval – This is for clarity purposes to ensure that users are aware that the confidence intervals are based solely off the percentages.

Changes to the column names from the previous release can be summarised below:

Previous column name	New column name
Screening LCI	Screening percentage LCI
Screening UCI	Screening percentage UCI
Two Week Wait LCI	Two Week Wait percentage LCI
Two Week Wait UCI	Two Week Wait percentage UCI
GP Referral LCI	GP Referral percentage LCI
GP Referral UCI	GP Referral percentage UCI
Other Outpatient LCI	Other Outpatient percentage LCI
Other Outpatient UCI	Other Outpatient percentage UCI
Inpatient Elective LCI	Inpatient Elective percentage LCI
Inpatient Elective UCI	Inpatient Elective percentage UCI
Emergency Presentation LCI	Emergency Presentation percentage LCI
Emergency Presentation UCI	Emergency Presentation percentage UCI
DCO LCI	DCO percentage LCI
DCO UCI	DCO percentage UCI
Unknown Route LCI	Unknown Route percentage LCI
Unknown Route UCI	Unknown Route percentage UCI
Route not classified LCI	Route not classified percentage LCI
Route not classified UCI	Route not classified percentage UCI

Any questions, please contact charlotte.eversfield@phe.gov.uk, sally.vernon@phe.gov.uk, anna.cooper@phe.gov.uk, sam.winters@phe.gov.uk.