Get Data Out technical documentation
Routes to Diagnosis

GDO 0019, 2020-03-05

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:

- AV2017.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
- AV2016.ROUTES2016 – This was used to select the Route to Diagnosis, linked to the cohort via tumourid. AV2016 was used in this instance as 2017 data were not available. This dataset was created following the Routes to Diagnosis 2006 to 2016 technical document.

2. Cohort

The cohort is created in line with the Get Data Out (GDO) incidence for each cancer site, which in turn are generated from the standard restrictions in the counting cases SOP. The cohort is created from AV2017.at_tumour_england. To be included in the cohort the following apply:

- Status of registration is final
- Country code is E
- Sex is either male or female
- The deduplication flag is set to 1

Cases not meeting the above 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 ROUTES2016 table was run on a cohort built from AV2016 for 2016. The 2013-2016 cohort from AV2017 includes some cases registered late that were not included in the ROUTES2016 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 existis, 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 2017 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’.

Any questions, please contact sam.winters@phe.gov.uk.