Public Health England

Protecting and improving the nation's health

Rapid Cancer Registration Dataset: ODR Guidance

Effective date: November 2020

This document gives a brief overview of the Rapid Cancer Registration Dataset to allow prospective ODR applicants to assess whether it is suitable to their needs.

What is the dataset?

The data set contains proxy tumour registrations and some associated events on the cancer patient pathway (e.g., surgery, radiotherapy and chemotherapy). These registrations are defined using an algorithm that approximates the cancer registration process using the datasets that the National Disease Registration Service has most rapid access to. It covers the period 2018-01-01 to the most recently available data, updated monthly by the National Disease Registration Service. There is some variation by data type but, on the whole, the dataset covers activity up to four to five months prior to the current date. The dataset overlaps with the gold standard cancer registration dataset for tumours diagnosed in 2018.

What purposes can the dataset be used for?

This dataset does not aim to replace the standard cancer registration data (and associated linked data), which is the preferred dataset to define diagnostics cohorts and allow linkage to other clinical activity data, and for detailed epidemiological research. The rapid cancer registration dataset should only be requested where the normal cancer data does not yet cover the period of activity of interest, and where the analytical question can be adequately answered despite the limitations of the data (see below). The analytical questions most likely to be adequately answered using this dataset are those concerning healthcare resource utilisation queries regarding the COVID-19 pandemic.

Structure and linkage

The dataset is formed of two tables (see [online documentation](#) complete field list). The RAPID_TUMOUR table holds basic data about the proxy tumours generated by the rapid registration process such as diagnosis date and cancer type: each row represents a proxy tumour diagnosis.

The RAPID_PATHWAY table holds event-based data, for example, information on the delivery of cancer treatments or the 'first seen' hospital attendance following a Two

1

Week Wait referral: each row represents one specific event on the cancer pathway. Most requests for rapid registration data are likely to require pathway data only for patients included in the RAPID_TUMOUR table. However, the RAPID_PATHWAY table holds data both on the patients included in the RAPID_TUMOUR table, and also for activity occurring in the data period related to other patients. Where the data in RAPID_PATHWAY relates to patients not included in RAPID_TUMOUR, this could be because:

- the patient was diagnosed prior to 2018 (in which case, tumour information is held in the normal cancer registration dataset),
- the rapid registration process has missed a cancer diagnosis,
- the patient has received radiotherapy or systemic anti-cancer therapy for a non-cancer diagnosis,
- the patient's HES record has an incorrect cancer code included in the diagnosis fields.

Tumours captured in the RAPID_TUMOUR table can be linked to 'standard' CWT, HES, SACT and RTDS datasets at patient level via NHSnumber. This allows RAPID_TUMOUR to be used for case finding while allowing the full richness of these secondary datasets to be utilised. Data availability and the completeness of these datasets should be carefully discussed with the NCRAS where the most recently available data is required.

Gold standard registrations are available within the RAPID_PATHWAY table (events 18 and 9 for final and provisional registrations respectively). This may allow project-specific assessment of the quality of the proxy tumours if the particular need is outside the scope of the generic data quality assessments available for the Rapid Cancer Registration Dataset. Only summary information is contained within these events but can be supplemented by the full cancer registration dataset if needed.

Some requests of the Rapid Cancer Registration Dataset may require repeated offloads on a regular timescale. Where this is the case it is expected that the specification of the data offload will remain constant over the course of the offload period. Re-specification of the dataset may require an additional application through the Office for Data Release and the ODR cost recovery policy will be applied.

Data quality and caveats
The dataset is explicitly built as a *rapid proxy* for cancer registration data. There are several known quality and completeness issues. These are more fully documented in a Data Quality document. In brief though, it is known by comparison to historic gold standard registration data that approximately 5% of the proxy tumours in the dataset

UNCONTROLLED FOLLOWING DISSEMINATION

are incorrectly included in the dataset in that they do not correspond to a registrable cancer, were all data to be fully processed according to cancer registration rules. In addition, approximately 18% of real tumours that can be expected to be eventually registered are missing from the dataset. These missing tumours are disproportionately in older persons and particular cancer types – see the data quality document for more details. These caveats are likely to bias analyses and should be carefully considered when determining whether to request the rapid registration data, and also when interpreting results based on the data.

Applying for access

Access to the Rapid Cancer Registration Dataset is controlled by the Office for Data Release through ODR Approval.

To obtain a copy of the ODR's pre-application pack (including the data request form and guidelines which describe the mandatory and qualified requirements for your application) or to seek support through the Pre-application Support Service (PaSS) contact ODR@phe.gov.uk.