

Rapid cancer registration in England, created in response to the COVID pandemic

Carolynn Gildea, National Disease Registration Service, NHS England

ENCR Workshop 4: IT tools and novel AI approaches for cancer registration, 13 Nov 2023

Overview

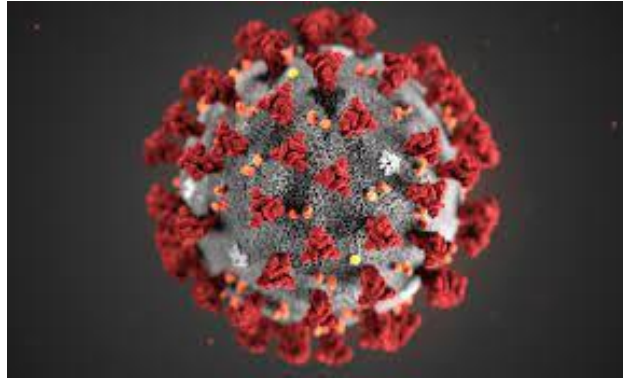
What is the Rapid Cancer Registration Data (RCRD)?

- Background – why it was developed
- What RCRD includes and how it is structured
- Development of RCRD
- Routine monitoring of data quality and reporting of results

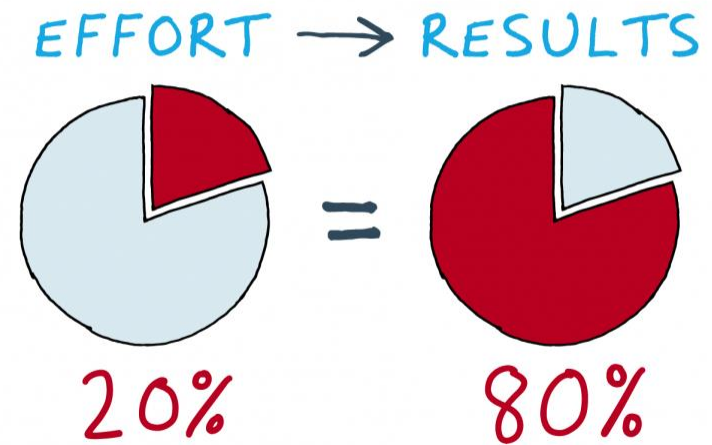
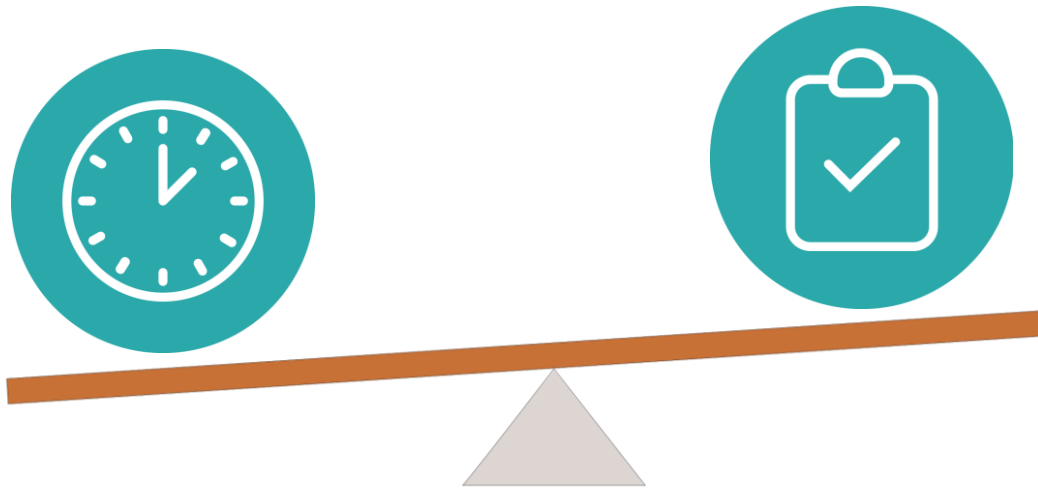
Find out more - conference poster:

Rapid cancer registration data: A response to the Covid-19 pandemic with long-term benefits

Why did we develop the Rapid Cancer Registration Data?

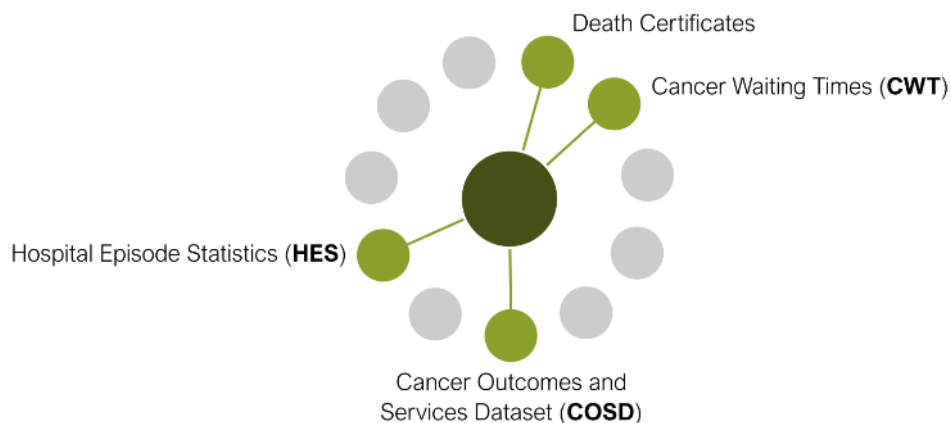


THE PARETO PRINCIPLE



What RCRD includes

Rapid Registration (**RCRD**)



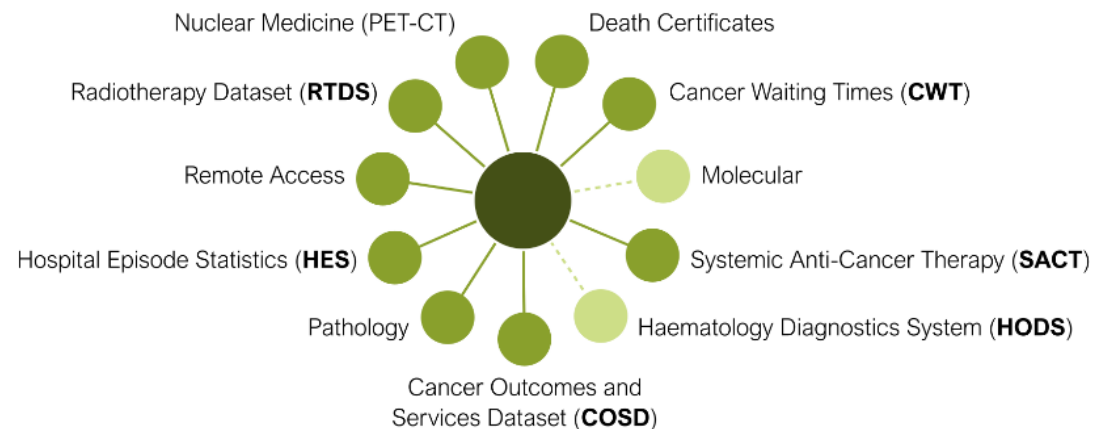
Malignant cancers (ICD-10 C00-C97) and some in-situ or benign tumours for bladder, breast and brain

Available 4 months post-diagnosis: rapid monitoring of case numbers

Staging for 13 cancer types

Available from January 2018

National Cancer Registration Data (**NCRD**), 'Gold Standard' data



All registerable conditions, ICD-10 C00-C97, D00-D48

Available 21 months post-diagnosis: full reporting, all clinical detail

Staging for all stageable cancers

Available from January 1995 (in ICD-10)

How RCRD is structured

RAPID_TUMOUR

TUMOUR_AVPID
INDIVIDUALID
PATIENTID
NHSNUMBER

DIAGNOSIS DATE
BASIS OF DIAGNOSIS
TUMOUR SITE
TUMOUR MORPHOLOGY
ROUTE TO DIAGNOSIS
CHARLSON COMORBIDITY
STAGE

BIRTH DATE
GENDER
POSTCODE
SURNAME
FORENAME
ETHNICITY



ID fields

Tumour fields

Patient fields

“What, where & when”

RAPID_PATHWAY

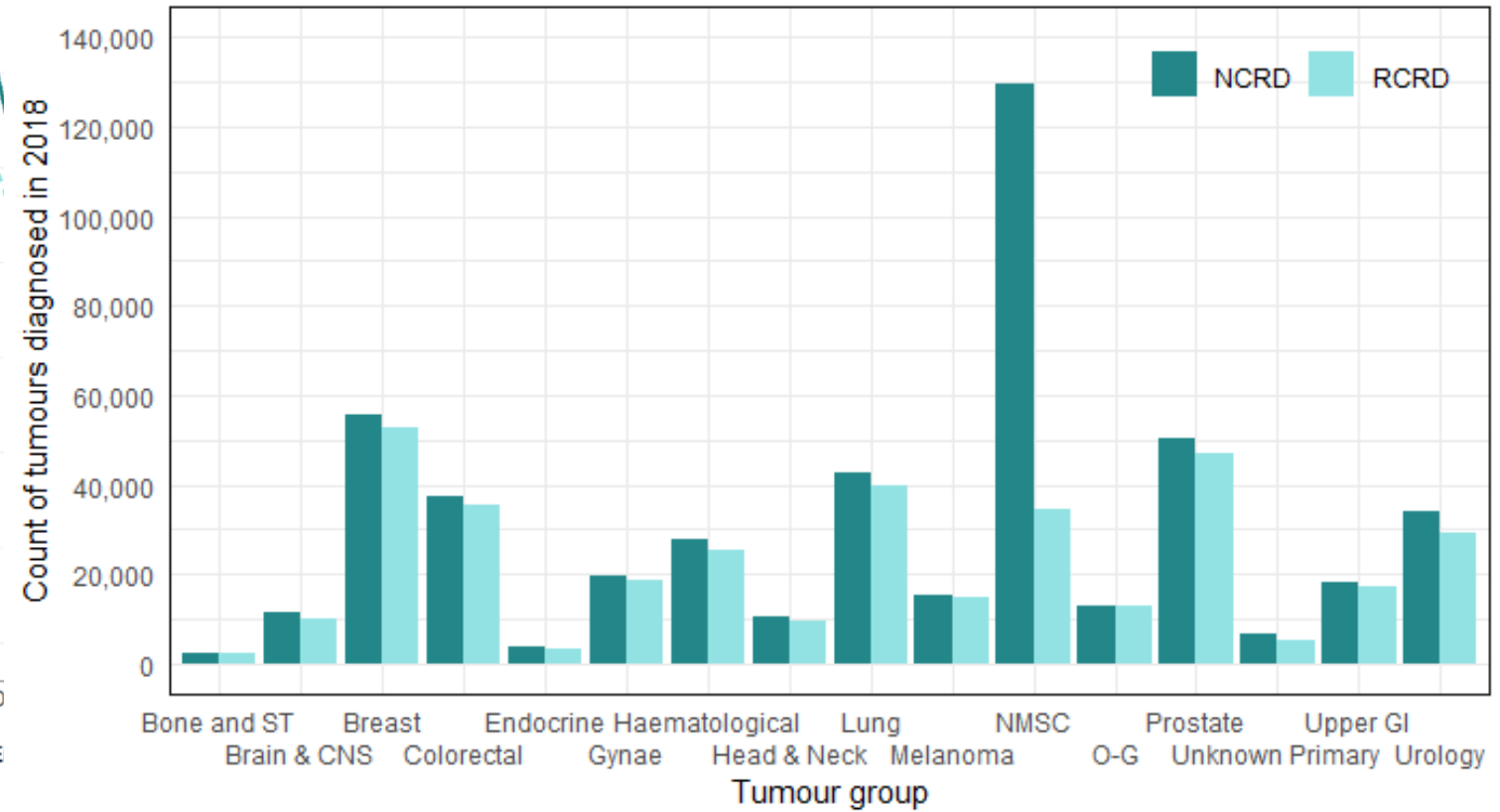
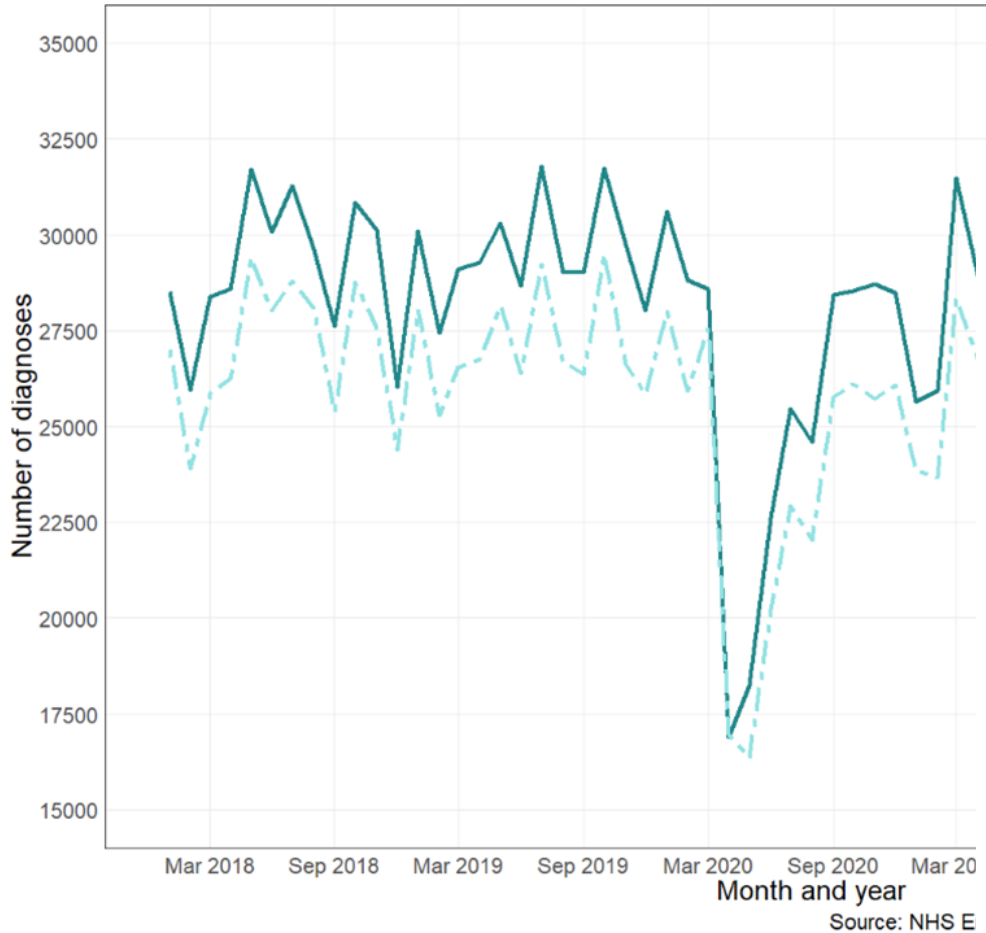
AVPID
INDIVIDUALID
PATIENTID
NHSNUMBER
SOURCE_TABLE
SOURCE ID

EVENT_TYPE
EVENT_PROPERTY_1
EVENT_PROPERTY_2
EVENT_PROPERTY_3
EVENT_DATE
EVENT_END
TRUST_CODE



34 event types from
several data sources

RCRD provides a reasonable proxy for NCRD



CNS: Central Nervous System; GI: Gastrointestinal;
 NMSC: Non-Melanoma Skin Cancer; O-G: Oesophago-gastric; ST: Soft Tissue
 Source: NHS England, National Disease Registration Service (CAS2310)

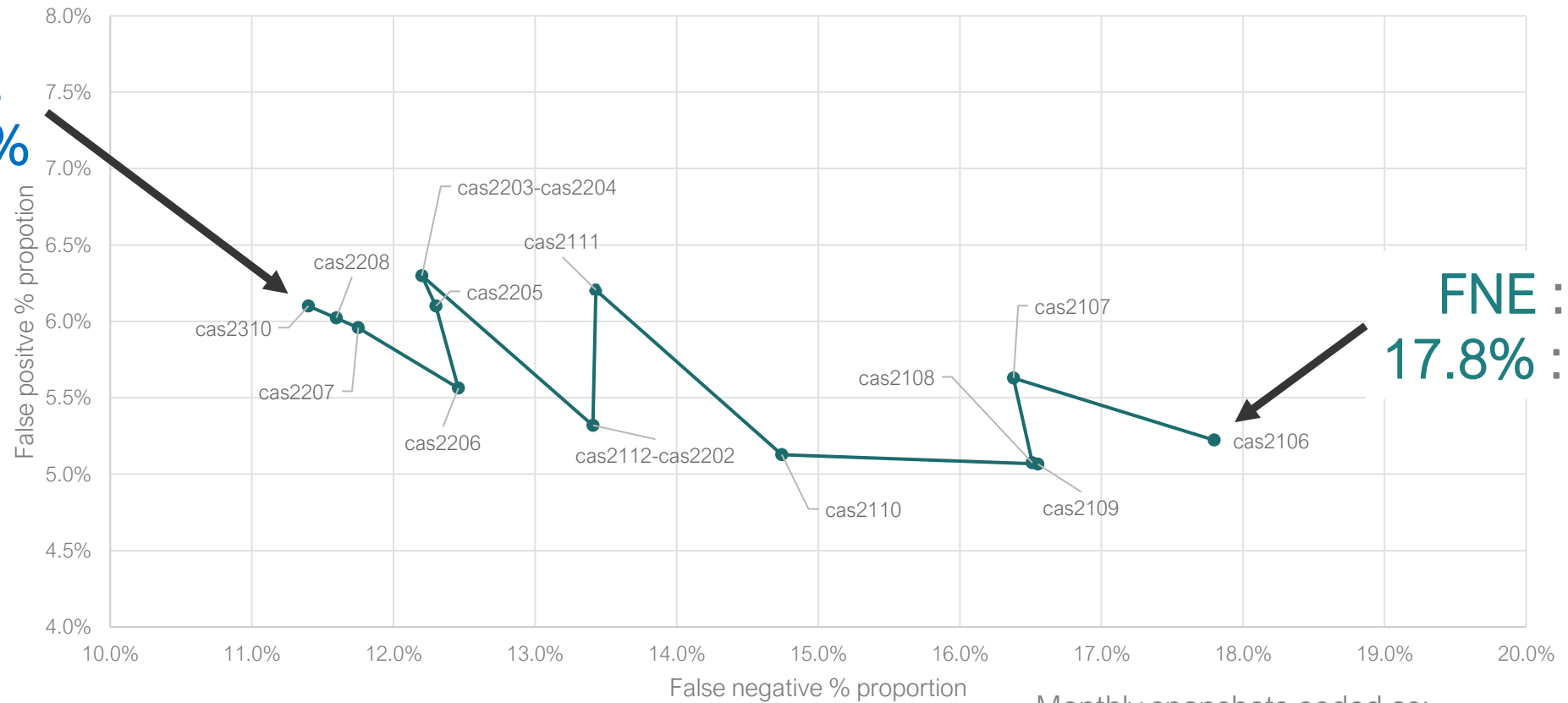
Defining quality – False Positives, False Negatives

- Two types of errors:
 - **False negative (FNE)** – real registration but no proxy registration (“*missing data*”) **11.4%**
 - **False positive (FPE)** – no real registration but proxy registration (“*bad data*”) **6.1%**
- We know some details about these errors:
 - **False negative:** Patterns with age and cancer type, related to clinical basis of diagnosis
 - **False positive:** Mixture of reasons e.g. difficult sites for ICD-10 codes; initial cancer record is changed later

Improving data quality – FNE : FPE

RCRD False Negative and False Positive error proportions by snapshot

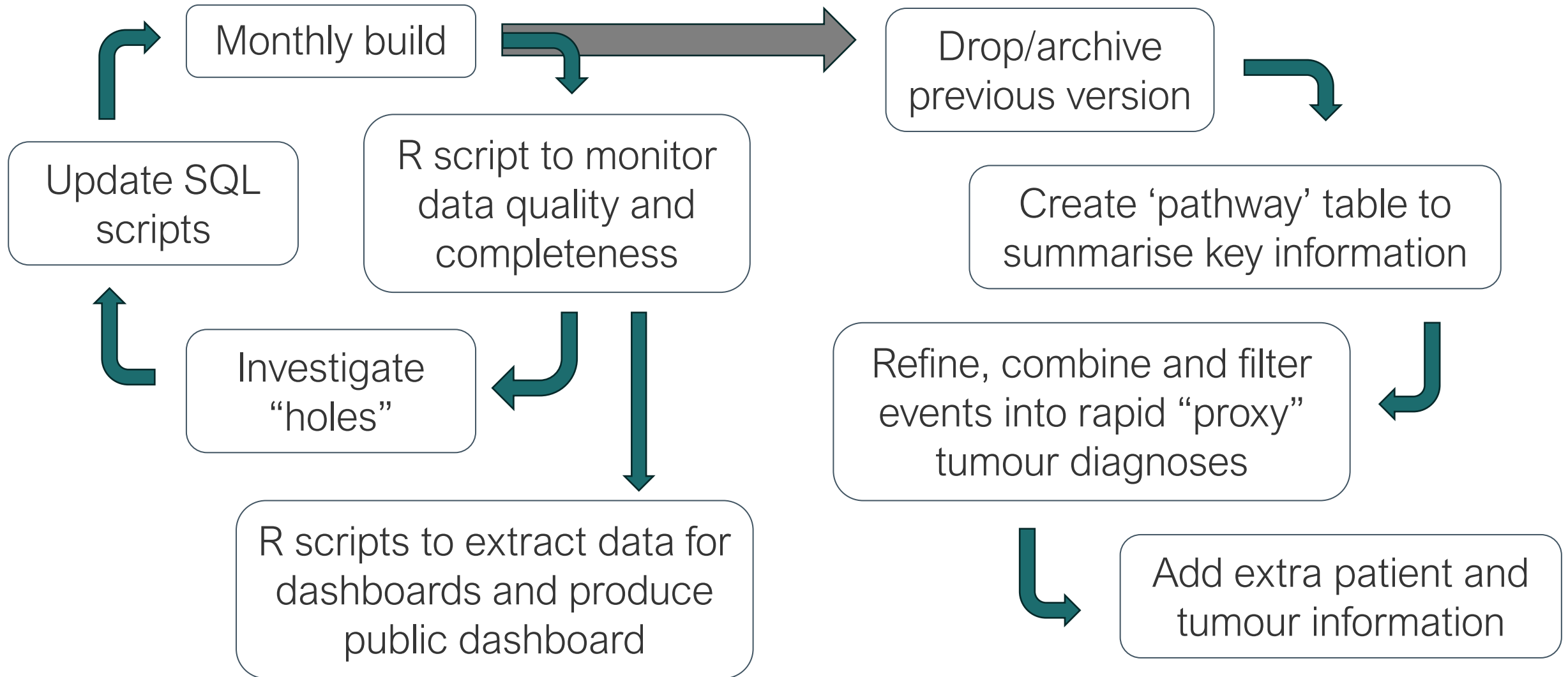
FNE : FPE
11.4% : 6.1%



FNE : FPE
17.8% : 5.2%

Monthly snapshots coded as:
casYYMM

Development cycle and monthly build process



Monthly reporting: Data quality and completeness

Rapid Cancer Registration Dataset: data at 7th October (CAS2310)

The National Cancer Registration and Analysis Service (NCRAS) has developed an algorithmically generated Rapid Cancer Registration Dataset (RCRD) using the standard administrative datasets which flow rapidly into NHS England (NHSE) and are incorporated into the Cancer Analysis System (CAS) of NCRAS. The data takes the form of a series of significant events that occur to each patient as they proceed through the diagnostic and then therapeutic parts of the cancer pathway, and is available at approximately 4-5 months behind real time. The RCRD is shallower and narrower than the full NCRAS cancer registration dataset; it should be used and interpreted with reference to the caveats outlined within this document.

Main findings

This document outlines the main features of the data to be aware of when interpreting the Rapid Cancer Registration Dataset:

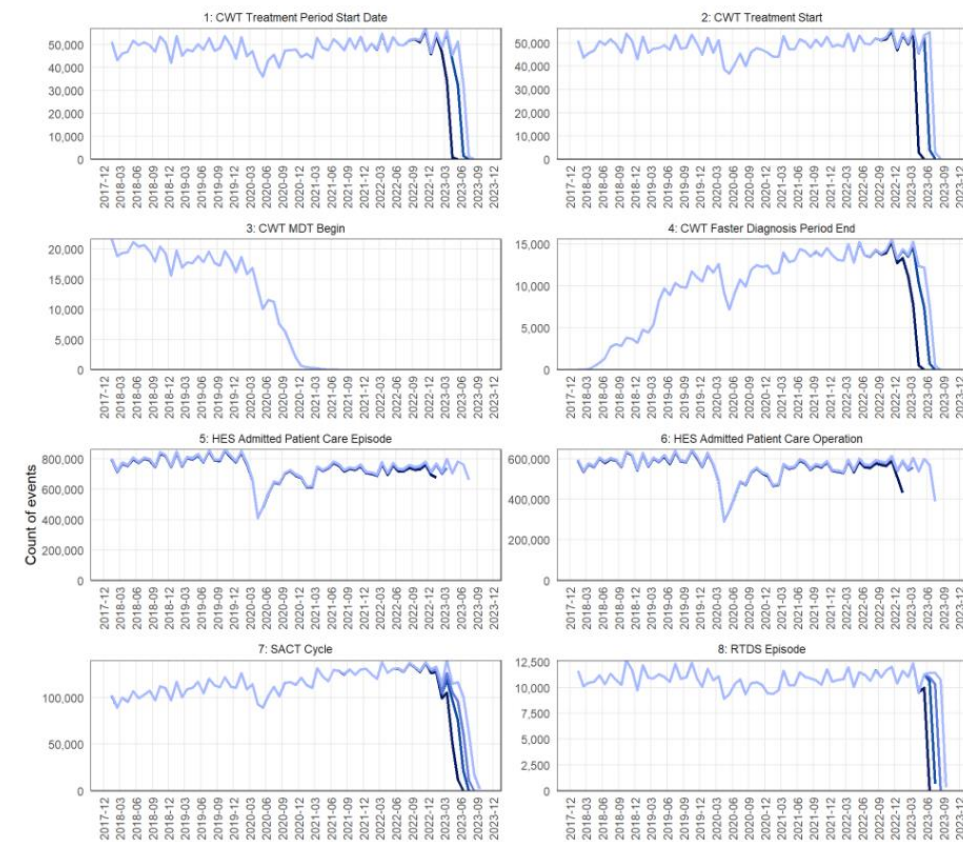
- Across all cancers types included approximately 11.4% of cases are missing and 6.1% of cases are included erroneously or with incorrect cancer type or diagnosis date (when compared to 'Gold Standard' registration data for 2018 data).
- These figures vary strongly with cancer site. Broadly, more common cancers (particularly breast and prostate cancer) perform best and less common cancers (particularly bone and soft tissue and cancers of unknown primary) perform worst.
- Non-melanoma skin cancer (ICD-10 C44) tumours are excluded from the majority of data shown (Figure 3 onwards). Carcinoma of the cervix (ICD-10 D06) is excluded from all data presented.
- There are more missing tumours in those aged over 70 compared to younger age groups.
- Other factors that reduce data completeness include the patient's route to diagnosis, mortality within 30 days or diagnosis, and the presence of multiple cancers.
- Usable data is available approximately 4-5 months after diagnosis or other clinical activity occurs.
- Data on cancer stage group at diagnosis is available for a number of common tumour types, although completeness is lower than that for the Gold Standard registration data. Where data is available it generally agrees with the Gold Standard stage group in 80-90% of tumours.

The dataset includes Rapid Cancer Registrations from January 2018 to the most recently available data (at the date specified in the title to this document), plus additional event data for the same period.

Contents

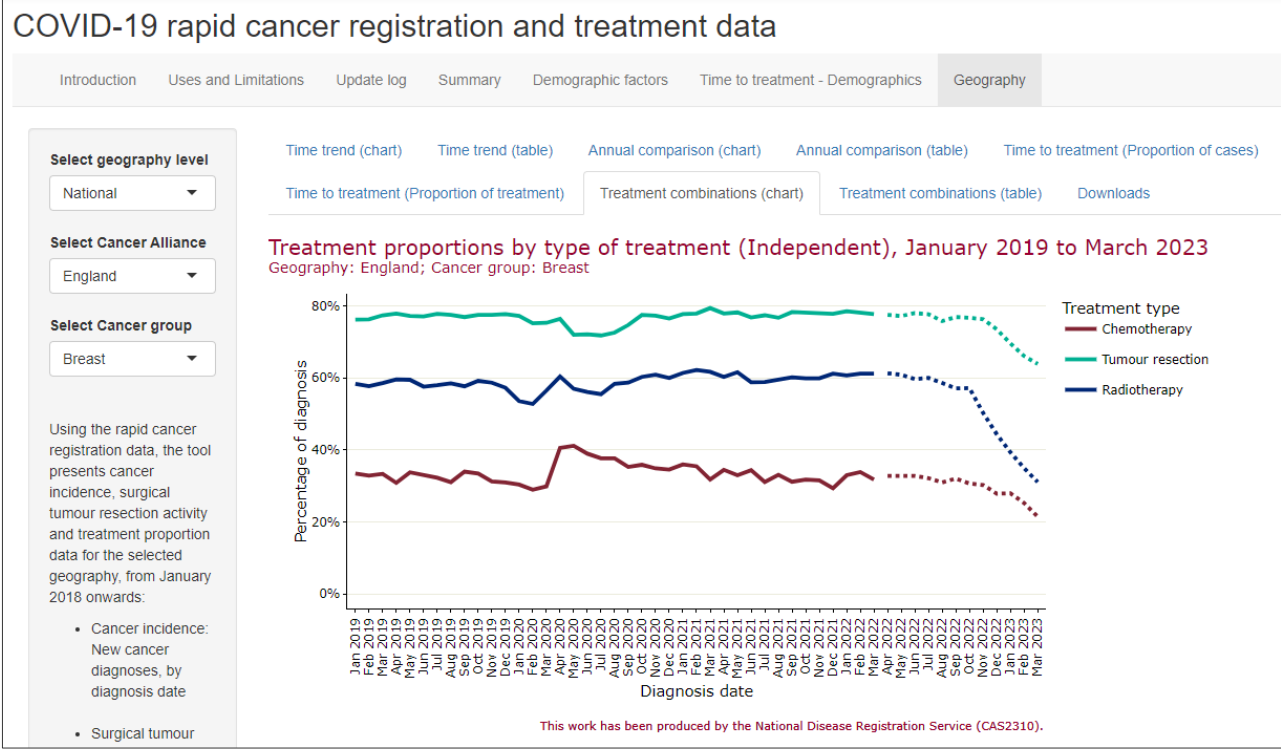
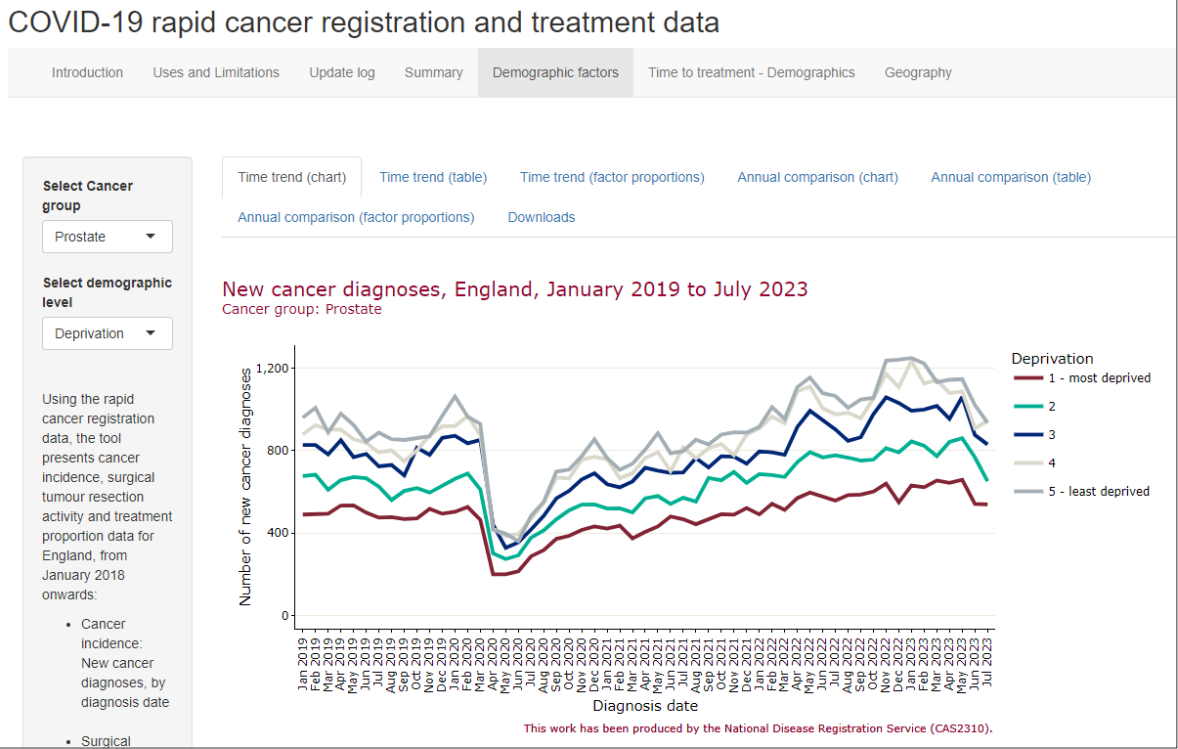
- Summary
- Methodology
 - Proxy registration events (Rapid Registrations)
- Data structures

Figure 14: Population of data items to CAS snapshot



NDRS website: digital.nhs.uk/ndrs/data/data-sets/rcrd#resources

Monthly reporting: Data dashboards



Cancer Data: cancerdata.nhs.uk/covid-19/rcrd

Concluding remarks

Progress to date:

- New dataset which is a reasonable proxy for cancer registrations
- Available around 4-months after diagnosis
- Includes 'key' patient and tumour characteristics
- Regularly used for service monitoring
- Although there are some biases, so it is not suitable for most epidemiological research

Next steps:

- Continue to develop the algorithm, to improve errors
- Look to extend data items included, quality and completeness (e.g. stage)
- Provide guidance about limitations

Many thanks to:

Sean McPhail

Tom Bacon

Jackie Charman

Cong Chen

Sophie Jose

Ravneet Sandhu

Sally Vernon

Kirstin Roberts

Hanhua Liu

Peter Jones

Any Questions?

Carolynn Gildea

Principal Analyst

carolynn.gildea@nhs.net

This work uses data that has been provided by patients and collected by the NHS as part of their care and support. The data are collated, maintained and quality assured by the National Disease Registration Service, which is part of NHS England.