Using and understanding cancer data

There is an increasing interest in collecting and using cancer data across the NHS. However, it can be difficult to understand what data are available, how they are collected, how they are used and how users can access data. In addition, there have been a number of changes to organisational structures supporting data collection

Below is a brief guide to cancer data in the UK (primarily in England) with a focus on the clinical rather than research perspective.

Data sources

Hospital Episode Statistics (HES): These data cover information on all hospital admissions, with coded data on diagnoses and procedures. They also covers outpatient attendances, but these often do not include any data on diagnosis or procedures.

Radiotherapy Dataset (RTDS): This is generated as a direct extract from the radiotherapy department treatment database, and so can contain very detailed information, including time of each fraction delivered. However, the clinical interpretation of these data is not always straightforward, especially for complex radiotherapy, where it can be difficult to understand the total dose delivered.

Systemic Anti-cancer Therapy Dataset (SACT): SACT data are collected directly from electronic prescribing systems. The dataset contains information on disease, morphology (i.e. histological diagnosis), performance status, regimen, drugs, doses and dates.

Mortality Data: The Office for National Statistics provides data on deaths, including dates and causes of death. Date of death information is available through the NHS Spine.

Diagnostic Imaging Dataset: This dataset contains a record (extracted from Radiology Information Systems (RIS)) of every imaging investigation undertaken, covering the date of the investigation, imaging modality and the area of the body imaged (but not the report).

National Cancer Registration Service (NCRS): This previously comprised 8 regional cancer registries, but there is now a unified service run by Public Health England. This contains data on every cancer tumour in England, including diagnosis, histology and some treatment data.

Clinical Outcomes and Services Dataset (COSD): This contains a feed from every hospital in England about every cancer patient, including a core for all patients (demographics, performance status) and disease specific elements (tumour grade, ER/PR and HER-2 for breast cancer, for example).

Cancer Waiting Times (CWT): These include data on referrals and treatment dates, and allow measurement against 14, 62 and 31 day standards.

How data are collected

Data are extracted via a variety of mechanisms. Cancer registry staff actively look for new cancers and receive a direct feed from all pathology departments; the RTDS and SACT Dataset are designed to flow from the electronic systems used in delivering radiotherapy and systemic therapy treatments.

COSD is generated from multi-disciplinary team meetings (MDTMs), as long as a suitable electronic MDT system is used. These data will flow directly from the computer system to the NCRS.

How to improve the quality of your data

The biggest complaints about cancer data are its quality and the uses to which they are put.

There are a few steps that can be undertaken to improve data quality:

1. Make sure you find all the patients
2. Look at the data regularly
3. Fix the holes
4. Look at the national datasets that you contribute to
5. Talk to your local NCRS data quality improvement officer.

Point 1 is probably the key element - without knowing which patients you have, it is impossible to improve data.

The key to 2 & 3 is to produce some form of regular feedback which highlights missing data. For example, a weekly spreadsheet, emailed to each MDT lead, can show data completeness against key metrics. This then enables identification of which patients are missing data, and enables gaps to be filled.

With regard to national datasets, there are now online portals for centres to review their data for COSD, RTDS and SACT. The portals allow centres to look at data completeness against key fields. You can access the data here: https://nww.cancerstats.nhs.uk (please note that this is only accessible from an NHS computer and that signup is required)

For centres identifying problems with data contained in national datasets, you are advised to discuss this with your local data improvement officer. Each region has a dedicated registry officer who works with the centres in their area to improve data capture.

Using data constructively

As data quality improves, the data can become much more clinically useful, and clinical engagement tends to drive improvements in data quality.

For example, one NHS Trust has used external NCRS quality-assurance of data to show that the estimated number of brain tumours was a significant under-estimate, and thus changes proposed to service arrangements (including closing units) were not sensible. This benefited from NCRS involvement, as the data collected were externally quality assured, and resulted in a significant improvement in the amount and quality of brain tumour data in the region, including a near-doubling in the apparent number of patients seen.

On a simpler level, clinicians can use HES data to obtain counts and numbers of patients under their care, or the number of patients having certain procedures which can then be used for revalidation purposes. Alternatively, clinicians can use data on dates of death, cross-linked with other data, to show - for example - 30-day mortality after chemotherapy.

The devolved nations

Most of the data described above are collected by Public Health England (PHE). The devolved nations have different data collection systems. For example, Scotland uses a different hospital dataset, and does not yet collect chemotherapy and radiotherapy data at a

national level. However, there are moves towards re-integrating some of the data collection systems, especially for radiotherapy and chemotherapy.

This is a simple introduction to cancer data. If you have more questions, or need to access data, please contact your local NCRS data liaison officer or your centre's IT department.

Please contact the author of this guide, RCR's Clinical Oncology Information and IT Resources Lead, Dr Matt Williams, with any questions about cancer data, or suggestions for correction, improvement and extensions to this document.