Data Resource Profile

Data Resource Profile: National Cancer Registration Dataset in England

Katherine E Henson,* Lucy Elliss-Brookes, Victoria H Coupland, Elsita Payne, Sally Vernon, Brian Rous and Jem Rashbass

National Disease Registration, National Cancer Registration and Analysis Service, Public Health England, London, UK

*Corresponding author. National Cancer Registration and Analysis Service, Public Health England, Wellington House, 133-155 Waterloo Rd, Lambeth, London SE1 8UG, UK. E-mail: katherine.henson@phe.gov.uk

Data resource basics

National Cancer Registration and Analysis Service

The National Cancer Registration and Analysis Service (NCRAS), part of Public Health England (PHE), is the population-based cancer registry for England. It collects, quality assures and analyses data on all people living in England who are diagnosed with malignant and pre-malignant neoplasms, with national coverage since 1971. It produces the national cancer registration dataset for England. The primary role of NCRAS is to provide near real-time, cost-effective, comprehensive data collection and quality assurance over the entire cancer care pathway. To achieve this, it receives data from across the National Health Service (NHS).

There are around 300 000 malignant tumours[1] diagnosed each year in England, for which NCRAS receives about 25 million records (Figure 1). These records are submitted by 162 health care providers, which incorporate over 1700 multidisciplinary teams (MDTs). MDTs are groups of professionals with expertise in different disciplines, whose remit is to assess relevant information about the patient and their cancer and make diagnosis and treatment recommendations; they are fundamental to cancer services in England.

Registration model

NCRAS employs an event-based registration model. Data on each primary tumour are sent from multiple sources to NCRAS, and either registered as a new cancer or linked to an existing cancer registration as a new event. A wide range of information is received and used to develop a rich data resource over time. The data are accessed through the Cancer Analysis System (CAS), which creates a monthly snapshot of data for analysis. A provisional quarterly snapshot is generated to facilitate timely data release, and then a designated annual snapshot when each calendar year of registrations is completed. The annual snapshot is used to produce national statistics.

Protection of cancer registration data

The provisions of Section 251 of the NHS Act 2006 provide the legal basis to Public Health England for collecting patient-level data on cancer patients for specified purposes, without consent. This is reviewed annually by the Confidentiality Advisory Group of the Health Research Authority.[2] Strict technical and contractual controls are put in place to prevent unauthorized access and use of the data, with staff undergoing regular training on data protection and information governance.

Data collected

The data collected by NCRAS comes from several sources including: multidisciplinary team meetings, pathology reports, molecular testing results, treatment records, hospital activity
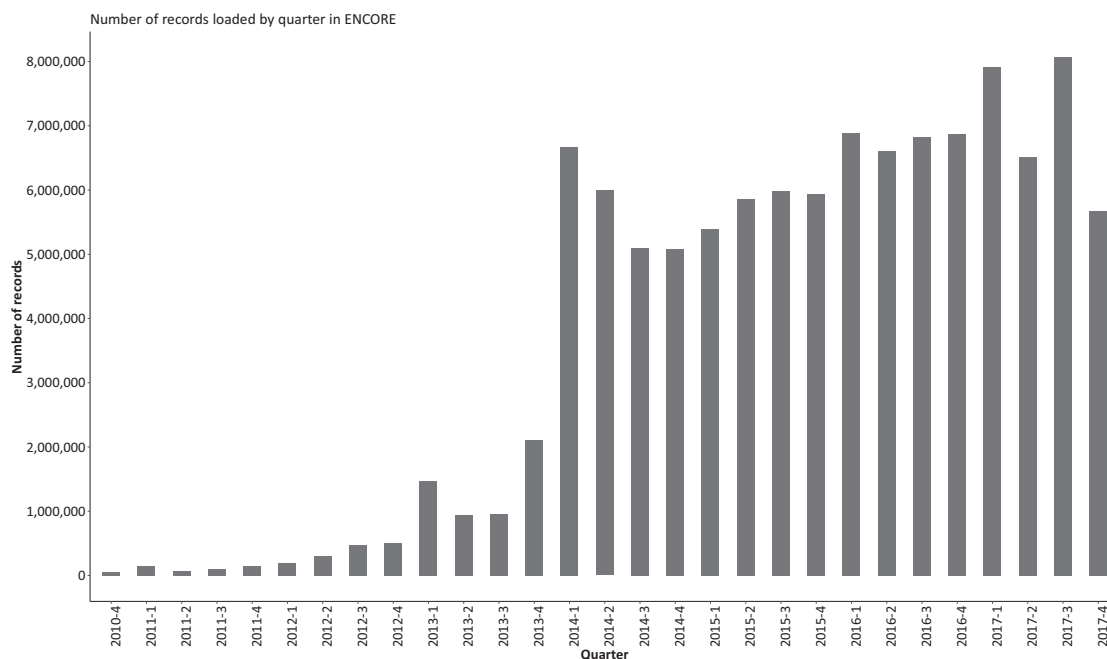
Figure 1. Number of records loaded (processed) by quarter by the National Cancer Registration and Analysis Service in ENCORE, as of December 2017.

records,[3] hospital Patient Administration Systems, and operational standards such as cancer waiting times.[4] The cause and date of death for deceased patients are supplied by the Office for National Statistics.[5] The primary patient identifier is the NHS number (a unique identifier used throughout the health care system in England), but date of birth, full name and address are also used for patient identification and linkage.

Registration is framed at the tumour level. Data submitted are reviewed by skilled cancer registration officers (CROs) with the assistance of some automated tools for data linkage and de-duplication of identical data sources. Review includes manual extraction of data from text-based pathology reports. CROs require detailed knowledge of cancer biology, coding and terminology. As cancer care is often delivered at different hospitals, information may relate to patient episodes at many organizations, which can introduce inconsistencies in the incoming data. The CRO will examine all data sources and, if necessary, seek additional source information from primary or secondary care via correspondence or direct interrogation of hospital radiology or electronic health records via remote secure access.

The data standard for the national cancer registration dataset

In January 2013, the Cancer Outcomes and Services Dataset (COSD) was introduced as the new national standard for reporting cancer in England.[6] The COSD data structure is extensive, containing 489 items in version 8,[7]
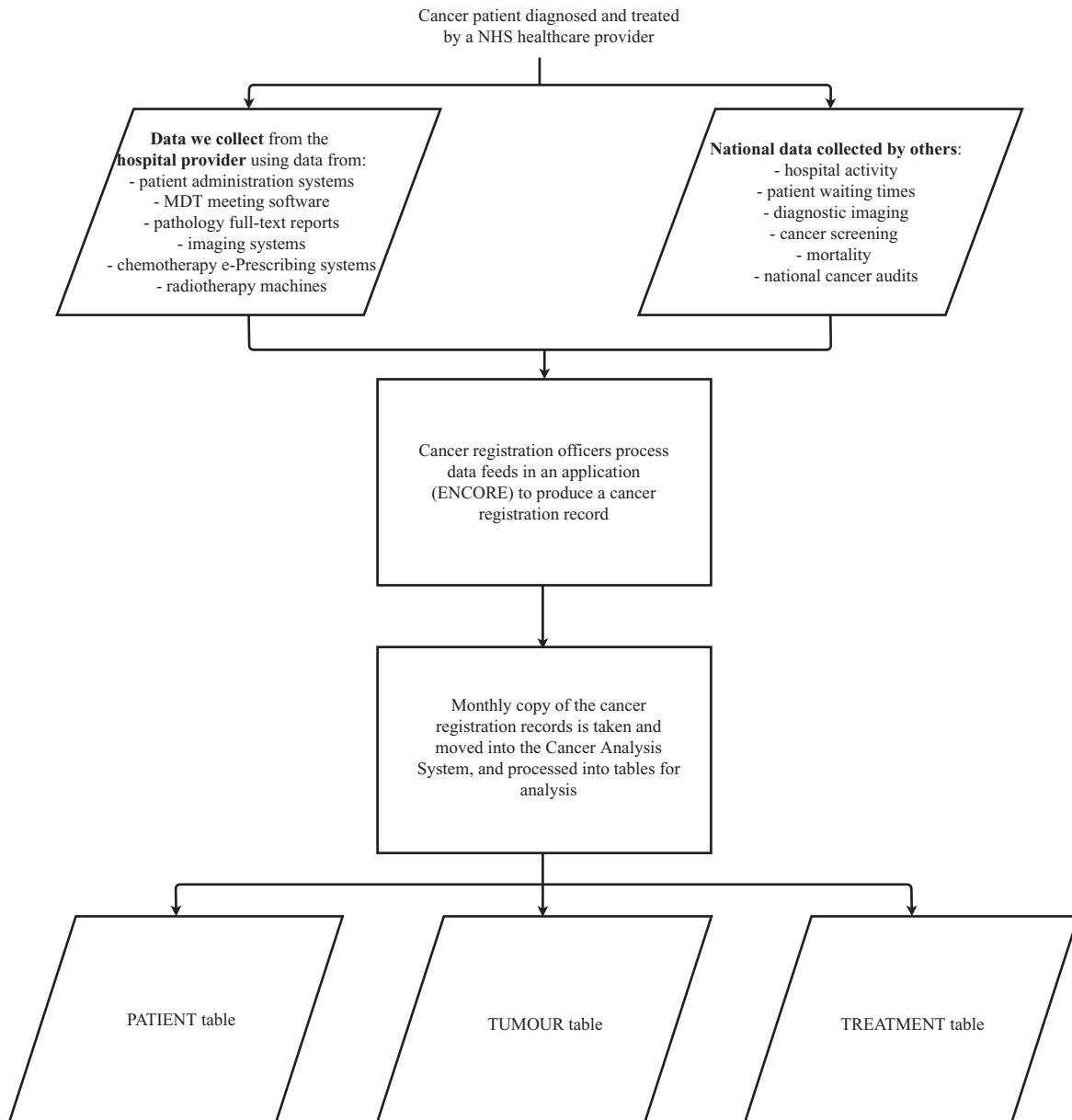
and covers clinical and pathological items. The data structure specifies the information to be sourced from a number of datasets, such as operational data on patient waiting times,[4] and mortality data.[5] The national cancer registration dataset includes a subset of COSD. The data standard undergoes periodic review to reflect changing clinical practice and incorporate new data sources.

New data sources

The cancer registration dataset is updated to reflect new iterations of COSD and additional data feeds. As clinical practice continues to evolve, and more detailed questions on cancer epidemiology and care postulated, new sources of relevant data are sought to both enrich and complement that collected by COSD. The most recent addition to the data collection undertaken by NCRAS is molecular data. Figure 2 provides an overview of the current data flows.

Quality assurance of cancer registration data

There are robust automated and manual quality control and assurance checks embedded into the registration process, both at the individual record level and for the set of registrations. As records are added to the system, input validation checks compare them with expected values and cross-validate against other records, for example if treatment occurred after death. Quality control checks are then undertaken during processing: each new registration is assessed by

Cancer patient diagnosed and treated
by a NHS healthcare provider

Data we collect from the
hospital provider using data from:
- patient administration systems
- MDT meeting software
- pathology full-text reports
- imaging systems
- chemotherapy e-Prescribing systems
- radiotherapy machines

National data collected by others:
- hospital activity
- patient waiting times
- diagnostic imaging
- cancer screening
- mortality
- national cancer audits

Cancer registration officers process
data feeds in an application
(ENCORE) to produce a cancer
registration record

Monthly copy of the cancer
registration records is taken and
moved into the Cancer Analysis
System, and processed into tables for
analysis

PATIENT table

TUMOUR table

TREATMENT table

ENCORE: English National Cancer Online Registration Environment
MDT: Multidisciplinary Team

Figure 2. Summary of the data flows from the point a patient was diagnosed with cancer to the analytical views of data in the Cancer Analysis System.

two CROs and any inconsistencies resolved. Finally, checks across all finalized records are performed before the quarterly sign-off of data for release, to find unprocessed records. Trend- and population-level checks are also performed.

Cancer registry data quality should be assessed in terms of completeness, validity, comparability and timeliness.[8,9] There are extensive published quantitative data which demonstrate quality, for example the performance indicators published by the United Kingdom and Ireland Association of Cancer Registries[10] and the Cancer Incidence in Five Continents (CI5) publications.[11] The CI5 publications report systematic

international indices of data quality, namely the proportion of microscopically confirmed cases and death certificate-only cases. English cancer registration data have been assessed for completeness and accuracy against randomized controlled trials data,[12] inpatient hospital data[13] and simulations,[14] with positive results. Validity, comparability and timeliness of the national cancer registration data are tested and published alongside the national statistics of both cancer incidence and survival.[15] Of note is that serious errors (for example an invalid date) have been found in fewer than 0.1% of registrations for more than 10 years.[15]

The percentage of microscopically verified cancers is an indicator of quality; for 2016, 85.3% of all malignant cancers were microscopically verified.[10] Cancers where the death certificate is the only information available indicate a failure to capture or process relevant health services information, and a high percentage indicates incomplete population coverage. The number of registrations where death certification is the only information available is reduced by actively looking for additional supporting data. Whereas registrations based solely on death certificate information represented more than 8% of cancers in the 1980s, due to the increasing availability of source documents to cancer registration staff this has been reduced to less than 1% of all registrations in 2016.

Derivation of key diagnostic information: incidence date and stage at diagnosis

Date of incidence is defined according to the European Network of Cancer Registries rules.[16] Key events from which NCRAS derives the date include: pathological verification; discussion at MDT; referral with suspicion of cancer; and diagnostic imaging. Staging information received includes: data from MDTs; pathology reports of biopsies and surgical treatment; imaging results; and post-mortems. This information is reviewed by CROs and collated into a consistent registry incidence date and a single registry-defined stage at diagnosis. The registry-defined stage combines all relevant information available to give a single anatomical stage at diagnosis, using the TNM (Tumour Node Metastases) classification system[17] where this is available for a tumour, or a site-specific staging system. In some circumstances the registry-defined stage differs from other stages recorded by a hospital. For example, where patients are managed at multiple hospitals, different components of the staging information may be received from each hospital. Another example is where the reporting pathologist may not have been aware of pre-operative chemotherapy or radiotherapy but NCRAS has access to this information. The importance of defining a comparable stage, particularly in the context of international comparisons, has been discussed in the literature.[18,19]

Data structure: registration and analysis

The cancer registration data are processed via a custom-built application called ENCORE (English National Cancer Online Registration Environment) and stored in an Oracle database. ENCORE is a live application. Monthly snapshots of the data are cloned into a separate Oracle database (*Cancer Analysis System*) for analysis and data release (Figure 2).

A summary of the key data items available in the cancer registration dataset for each cancer is in Table 1. Full details can be found in the published data dictionary[20] which is updated periodically. The cancer registration dataset comprises three main tables: the tumour table holds information about each primary tumour diagnosed in an individual; the patient table holds information about the individuals who are diagnosed with cancer; the treatment table holds the treatment received by the patient. Several data items in the tables are derived using algorithms and multiple data sources, for example the Charlson comorbidity index score[21] derived using hospital admission data.[3,22]

Data resource use

English cancer registration data are used nationally and internationally for a wide range of functions including public health, commissioning, evaluating clinical performance, health care services and research. Specific examples include tracking of cancer incidence rates and survival statistics[23]; measuring of outcomes for people diagnosed with cancer[24]; informing evidence-based decisions about NHS service provision and patient care[25]; and international comparison of the survival rates of patients diagnosed with cancer in England.[26]
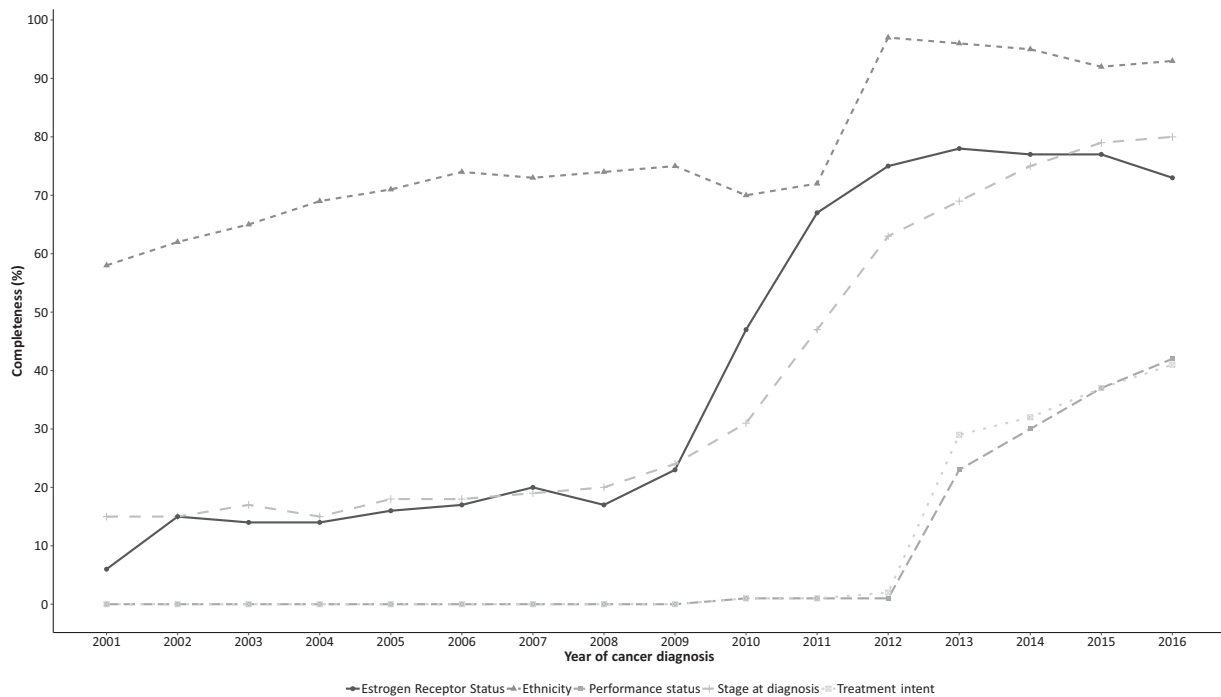
Improvements in data completeness and quality have allowed more extensive and useful analyses to be undertaken, as demonstrated by three key NCRAS publications. First, improved completeness of stage at cancer diagnosis (Figure 3) has permitted the publication of 1-year survival by stage at diagnosis[27] and quantified the survival deficit in those people diagnosed at a late stage. Second, categorizing the entire patient pathway using all available data has highlighted the variation in time spent in each part of the pathway.[28] Third, a pilot project has published granular data on less common cancers for use by patients and the public while protecting confidentiality.[29]

Since PHE was formed in 2013, there have been at least 260 publications based on English cancer registration data (as at 6 April 2018), excluding studies which used registration data to follow up randomized controlled trials. This is likely to be an underestimate, as a standard acknowledgement of cancer registration data has not been used consistently and identification of studies using the data has therefore been challenging.[30] Many publications use datasets which pre-date NCRAS, including the previous national dataset (National Cancer Data Repository) and regional registry data. Published papers demonstrate a wealth of studies aiming to improve outcomes for patients diagnosed with cancer; specific references are included in

Table 1. Summary of key data items available for each cancer registration

Patient	Tumour	Diagnosis	Treatment	Death
Patient identifier	Tumour identifier	Date of incidence	Event identifier	Date of death
NHS number	Site, morphology and behaviour of tumour	Basis of diagnosis	Type of treatment event (surgery/radiotherapy/chemotherapy)	Full coded causes of death from death certificate
Date of birth	Multifocal flag	Route to diagnosis	Date of event	Coded underlying cause of death
Sex	Tumour size	Health care provider at initial contact	Treatment health care provider	Location of death
Ethnicity	Stage: registry-derived stage at diagnosis and other stages	Health care provider at diagnosis	Indicator for whether patient in a clinical trial	Post-mortem
Postcode at diagnosis	Laterality	Date of MDT[a] meeting	Details of event (dependent on type of event)	
Comorbidity score (derived from linked hospital inpatient information)	Grade	Cancer care plan intent	Surgical information recorded using international coding system	
Performance status (at diagnosis)	Site-specific fields (e.g. Gleason grade for prostate cancer)	Record if patient was seen by a clinical nurse specialist	Type of imaging and site	
General Practice of the patient (at diagnosis)				
Deprivation (derived from postcode of residence at diagnosis)				

[a]MDT: Multidisciplinary team.

Figure 3. Trend in data completeness (%) from 2001 to 2016 of the recording of estrogen receptor status for breast cancer patients, self-reported ethnicity, the performance status at diagnosis for adults, the registry-derived stage at cancer diagnosis, and the intent of treatment.

the Supplementary material, available as Supplementary data at *IJE* online.

Strengths and limitations

National coverage

A key strength of the data collected and held by NCRAS is the complete coverage of all people diagnosed with cancer in England i.e. it is population-based. This enables the calculation of representative national epidemiological measures for all cancer types, including rarer types such as eye and heart. Population-based national statistics on cancer incidence are available from 1971.[23] However, some time series should be interpreted with care, as clinical and coding definitions of invasive cancer have changed over time and the list of specific cancers that are registered has reflected this. For example, there have been changes in which tumours are registered as invasive bladder cancer, and changes in coding practices were not consistently adopted by regional cancer registries, leading to misleading trends in reported incidence.[31]

Only NHS health care providers are mandated to submit data to NCRAS, and this includes NHS-funded activity undertaken in the private sector. It is estimated that 98–99% of hospital activity in England is funded by the NHS.[32] Although some private hospitals do also submit data, information on patients treated in the private sector will be incomplete.

Standardized coding

All registrations are coded to international standards. Since 2013, the site and morphology of the tumour are coded using the International Classification of Diseases for Oncology (ICD-O3) classification.[33] All tumours are coded with a specific site, morphology and behaviour classification code, which records sufficient granularity to map to other major coding classification systems (including ICD-10 for analysis)[34] and to classifications used specifically for childhood cancers[35] and teenage and young adult cancers.[36] Of 1 971 287 tumours diagnosed in England during 2013–16, 99.98% were registered using ICD-O3 and 99.96% have been successfully mapped to a valid ICD-10 topology and morphology code. This permits international comparisons. Regional variations existed in coding before 2012, particularly regarding treatment information, with differing inclusions of procedures such as open biopsies, immunotherapy and hormone therapy. Data sources available to registration staff were also greatly standardized by the national move to ENCORE from regional systems in 2012. Therefore, changes seen in data between 2012 and 2013 may not reflect clinical changes.

Timeliness

Final registrations are released approximately 1 year following the end of a diagnosis year. Following initial registration by a CRO, a 6-month period is allowed for treatment to occur, the health care provider to submit data and the data to be processed. The registration is then reviewed and finalized. This time lag results in improved data completeness and quality, but it prevents the publication of more timely data. Once each year of data is assessed as being complete, it is released to produce national statistics; however, CROs will continue to process any data received about new tumours, and thus the number of cancer registrations for previous years may increase slightly. It can take up to 5 years to reach 100% completeness.[15] This can occur because of late submission of data from health care providers, or can be due to seeking further data after initial notification via death certificate. These changes are relatively small: for example, between July and December 2017, each month there were on average 23 011 registrations for 2016 compared with 484 and 160 registrations for 2015 and 2014, respectively.

Treatment information

In addition to the treatment data directly processed by NCRAS, cancer registration data can be linked to a wealth of treatment datasets. Specifically these include data from e-prescribing systems for chemotherapy,[37] data from radiotherapy machines[38] and hospital activity data[3] for inpatient, outpatient and accident and emergency hospital admissions.

NCRAS processes all treatment information that occurs during the first 6 months following diagnosis. Data outside this period are collected and added to the tumour registration but may not necessarily be used to amend the finalized registration. Currently information on both the intent of treatment and the performance status is limited[39] (Figure 3).

Data completeness

In recent years, the completeness of specific data items such as stage and ethnicity has greatly improved. When new data items are added to COSD, the completeness is often low due to difficulty in data collection by health care providers, for example performance status (Figure 3). The NCRAS Data Liaison team provide direct support to hospital providers to improve their ability to collect data; this includes system procurement support, review of COSD data, clinical engagement and data reporting and validation.

The impact of opt-outs

The legal permission held by PHE does not require individual consent. Individuals may opt out and ask that their

health care information be removed from the National Cancer Registration Dataset. Very few (<1 in 10 000) patients have opted out and therefore the impact is minimal. Large numbers of opt-outs would impair the ability to produce meaningful statistics, particularly for rarer cancers, so NCRAS is working with major cancer charities to improve public and patient awareness of the value of population-based cancer registration.

Primary care information

Primary care information is available to link to cancer registration data for a subset of the population.[40] It is not available at the national level.

Patient outcomes

The linkage to mortality data allows the follow-up of patients to death for all causes of death. However, non-death outcomes, such as long-term treatment effects or quality of life assessments, are difficult to identify using registration data, particularly those that do not result in an inpatient hospital stay. Cancer recurrence and progression are currently poorly recorded in both hospital and registration data[41] due to a number of challenges, including inconsistent definitions and the time lag between initial presentation and the cancer recurrence, which can be many years after the original diagnosis. Data items for recurrence and progression have been included in COSD for primary diagnoses from 2014 onwards.

Risk factor information

Limited information is currently collected on lifestyle factors, such as smoking and dietary information. Body mass information can be ascertained for patients with chemotherapy recorded,[37] although data quality is variable.

De-duplication of cancer registration data

The migration of regional cancer registries into ENCORE was completed in 2012. Before 2012, if a patient was diagnosed with cancer or treated in more than one registry catchment area they may have been registered multiple times. These duplicates were resolved for the publication of national statistics, but the registrations would have remained in the regional registry data. The merging of regional data in ENCORE brought together all registrations into a single system including duplicates (of patients and tumours). Over 450 000 patient duplicates have been resolved on ENCORE across all years of data available (national coverage since 1971), and the data are fully de-duplicated from 2012 onwards. As an example, for years preceding this, around 1% of patients are flagged as duplicates (based on patient NHS number and date of birth). All duplicates are routinely excluded from analyses and data releases.

Data resource access

Access to cancer registration data can be granted to individuals who demonstrate that either there is a justified purpose for the data release, and that there is an appropriate legal basis with safeguards in place to protect the data, or the data release is deemed to be anonymous (e.g. aggregate data). Dependent on the request, ethical approval may be required. The access process is managed by the Office for Data Release. A cost may be associated with the data release, and this is agreed before work commences. Full details, including data dictionaries, are available on the website.[20] Pre-application advice can be sought from [odr@phe.gov.uk].

For cancer registration data releases via the PHE Office for Data Release, the following acknowledgements statement should be used: 'This project involves data derived from patient-level information collected by the NHS, as part of the care and support of cancer patients. The data are collated, maintained and quality assured by the National Cancer Registration and Analysis Service, which is part of Public Health England (PHE). Access to the data was facilitated by the PHE Office for Data Release'.

In addition to data releases, NCRAS supports collaborations with academic and other institutions. Enquiries should be directed to [NCRASenquiries@phe.gov.uk].

Conclusions

NCRAS registers all cases of cancer that occur in people living in England, resulting in a clinically rich data resource. Detailed clinical information on cancer across the full patient pathway enables the measurement of the diagnosis, treatment and outcome of all patients diagnosed with cancer. This population-based national data resource is critical to support service provision, clinical audit, commissioning, planning of services, public health and epidemiological research, all of which contribute to improved outcomes for people diagnosed with cancer.

Glossary

CI5: Cancer Incidence in Five Continents publications.

COSD: Cancer Outcomes and Services Dataset. This is the national standard for reporting cancer in England, and defines the data submitted to NCRAS by health care providers.

CRO: cancer registration officer. These officers manually assess the multiple data sources submitted to NCRAS to produce a cancer registration record for each diagnosis of cancer.

ENCORE: English National Cancer Online Registration Environment. This is a custom-built application used by the cancer registration officers to process all data submitted by health care providers to produce a cancer registration record.

ICD-10: International Statistical Classification of Diseases and Related Health Problems, 10th Revision. This is the international standard for classifying diseases and is published by the World Health Organization.

ICD-O3: International Classification of Diseases for Oncology, version 3. This is the international standard for classifying diseases in oncology.

NCRAS: National Cancer Registration and Analysis Service.

NHS: National Health Service.

MDT meeting: multidisciplinary team meeting. These are fundamental to cancer services in England and provide a multidisciplinary decision-making process for the diagnosis and treatment of cancer patients.

PHE: Public Health England.

TNM: Tumour Node Metastases. This is the internationally accepted standard for cancer staging, published in affiliation with the Union for International Cancer Control.

Profile in a nutshell

- NCRAS collects and quality assures detailed clinical information on all cancers that occur in people living in England. This national cancer registration data is used to support cancer epidemiology, public health, health care service monitoring and research. The expert input of highly trained cancer registration officers into the registration process ensures accurate patient, tumour and treatment information.

- Incidence data are available for cancers diagnosed since 1971. The coverage is national and population-based.

- Data collected as part of the health care of the patient are routinely submitted to NCRAS by providers.

- The data collected include: patient identifiers; patient-specific items (e.g. sex, ethnicity); cancer tumour-specific items (e.g. diagnosis date, cancer site, morphology); cancer site-specific items [e.g. estrogen receptor (ER) status, laterality]; stage at cancer diagnosis data [e.g. Tumour, Node, Metastasis (TNM), FIGO]; cancer treatment data (e.g. procedure code); death information (e.g. date and cause of death); provider information (e.g. hospital of diagnosis, GP practice code); and geography at diagnosis (based on patient's postcode of residence).

- NCRAS data are accessible though applications to the Office for Data Release. They are also used to enrich other data sources (such as the Clinical Practice Research Datalink) and generate national statistics.

Supplementary data

Supplementary data are available at *IJE* online.

Acknowledgements

References

1. National Cancer Registration and Analysis Service. *CancerData: Incidence*. https://www.cancerdata.nhs.uk/incidence/base_num bers (21 September 2018, date last accessed).

2. NHS Health Research Authority. *Section 251 and the Confidentiality Advisory Group (CAG)*. http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/

3. NHS Digital. *Hospital Episode Statistics*. http://content.digital.nhs.uk/hes (22 January 2018, date last accessed).

4. NHS Digital. *Cancer Waiting Times*. https://digital.nhs.uk/cancer-waiting-times (26 February 2018, date last accessed).

5. Office for National Statistics. *Mortality Data*. http://content.digital.nhs.uk/onsmortality (26 February 2018, date last accessed).

6. National Cancer Registration and Analysis Service. *Cancer Outcomes and Services Dataset*. http://content.digital.nhs.uk/isce/publication/dcb1521 (26 February 2018, date last accessed).

7. National Cancer Registration and Analysis Service. *Cancer Outcomes and Services Dataset, Version 8.0.1.* 2017. http://www.ncin.org.uk/view?rid=3451

8. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. *Eur J Cancer* 2009;**45**:747–55.

9. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. *Eur J Cancer* 2009;**45**:756–64.

10. United Kingdom and Ireland Association of Cancer Registries (UKIACR). *Performance Indicators*. 2018. http://www.ukiacr.org/kpis

11. Parkin DM, Ferlay J, Curado M-P *et al*. Fifty years of cancer incidence: CI5 I–IX. *Int J Cancer* 2010;**127**:2918–27.

12. Merriel SWD, Turner EL, Walsh E *et al*. Cross-sectional study evaluating data quality of the National Cancer Registration and Analysis Service (NCRAS) prostate cancer registry data using the Cluster randomised trial of PSA testing for Prostate cancer (CAP). *BMJ Open* 2017;**7**:e015994.

13. Møller H, Richards S, Hanchett N *et al*. Completeness of case ascertainment and survival time error in English cancer registries: impact on 1-year survival estimates. *Br J Cancer* 2011;**105**:170–76.

14. Woods LM, Coleman MP, Lawrence G, Rashbass J, Berrino F, Rachet B. Evidence against the proposition that "UK cancer survival statistics are misleading": simulation study with National Cancer Registry data. *BMJ* 2011;**342**:d3399.

15. Office for National Statistics. *Cancer registration statistics QMI*. 2016. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/cancerregistrationstatisticsqmi

16. Tyczynski JE, Parkin D. Standards and guidelines for cancer registration in Europe. *IARC Tech Publ* 2003;**40**:69–73.

17. Union for International Cancer Control. *TNM Classification of Malignant Tumours*. 8th edn. New York NY: Wiley, 2016.

18. Eden M, Rous BA, Rashbass J. Misinterpretation of the origins and composition of staging data and its impact on colorectal cancer survival. *Acta Oncol* 2014;**53**:845–46.

19. Walters S, Maringe C, Butler J, Brierley JD, Rachet B, Coleman MP. Comparability of stage data in cancer registries in six countries: lessons from the International Cancer Benchmarking Partnership. *Int J Cancer* 2013;**132**:676–85.

20. Public Health England. *Accessing PHE Data Through the Office for Data Release*. https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-phe-odr-and-accessing-data (26 February 2018, date last accessed).

21. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;**40**:373–83.

22. Elliss-Brookes L, McPhail S, Ives A *et al*. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. *Br J Cancer* 2012;**107**:1220–26.

23. Office for National Statistics. *Cancer Registration Statistics, England*. 2017. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancerregistrationstatisticscancerregistrationstatisticsengland (26 February 2018, date last accessed).

24. Office for National Statistics, National Cancer Registration and Analysis Service. *Geographic Patterns of Cancer Survival in England: Adults Diagnosed 2011 to 2015 and Followed up to 2016*. 2018. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/geographicpatternsofcancersurvivalinengland/adultsdiagnosed2011to2015andfollowedupto2016 (26 February 2018, date last accessed).

25. Royal College of Physicians. *National Lung Cancer Audit Annual Report 2017 (for the Audit Period 2016)*. London: Royal College of Physicians, 2018.

26. Allemani C, Matsuda T, Di Carlo V *et al*. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* 2018;**391**:1023–75.

27. National Cancer Registration and Analysis Service, Office for National Statistics. *Cancer Survival by Stage at Diagnosis for England (Experimental Statistics): Adults Diagnosed 2012, 2013 and 2014 and Followed up to 2015*. 2016. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancersurvivalbystageatdiagnosisforenglandexperimentalstatistics/adultsdiagnosed20122013and2014andfollowedupto2015 (26 February 2018, date last accessed).

28. National Cancer Registration and Analysis Service, Transforming Cancer Services Team for London. *Segmented Analysis of Prostate Cancer Pathway from Referral to Treatment: 2013–2015*. 2017. http://www.ncin.org.uk/view?rid=3544 (26 February 2018, date last accessed).

29. National Cancer Registration and Analysis Service. *Standard Output Tables: Routine Statistics are Made Available for a Standard Grouping of Cancer Patients in our Standard Output Tables*. 2017. https://www.cancerdata.nhs.uk/standardoutput (26 February 2018, date last accessed).

30. Rashbass J, Peake M. The evolution of cancer registration. *Eur J Cancer Care (Engl)* 2014;**23**:757–59.

31. Shah A, Rachet B, Mitry E, Cooper N, Brown CM, Coleman MP. Survival from bladder cancer in England and Wales up to 2001. *Br J Cancer* 2008;**99**: S86–89.

32. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol* 2017;**46**:1093–93i.

33. World Health Organization. *International Classification of Diseases for Oncology (ICD-O)*, First Revision. 3rd edn. Geneva: World Health Organization, 2013.

34. World Health Organization. *ICD-10: International Statistical Classification of Diseases and Related Health Problems*, 10th Revision. Geneva: World Health Organization, 2010.

35. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. *International Classification of Childhood Cancer*, third edition. *Cancer* 2005;**103**:1457–67.

36. Birch JM, Alston RD, Kelsey AM, Quinn MJ, Babb P, McNally RJ. Classification and incidence of cancers in adolescents and young adults in England 1979-1997. *Br J Cancer* 2002;**87**:1267–74.

37. National Cancer Registration and Analysis Service. *Systemic Anti-Cancer Therapy Dataset*. http://www.chemodataset.nhs.uk/home (26 February 2018, date last accessed).

38. National Cancer Registration and Analysis Service. *RadioTherapy DataSet*. http://www.ncin.org.uk/collecting_and_using_data/rtds (26 February 2018, date last accessed).

39. Henson KE, Fry A, Lyratzopoulos G, Peake M, Roberts KJ, McPhail S. Sociodemographic variation in the use of chemotherapy and radiotherapy in patients with stage IV lung, oesophageal, stomach and pancreatic cancer: evidence from population-based data in England during 2013-2014. *Br J Cancer* 2018;**118**:1382–90.

40. Herrett E, Gallagher AM, Bhaskaran K *et al*. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015;**44**:827–36.

41. NHS Digital. *OPCS Classification of Interventions and Procedures*. 2017. http://content.digital.nhs.uk/isce/publication/scci0084 (26 February 2018, date last accessed).