Data Resource Profile

Data Resource Profile: The Systemic Anti-Cancer Therapy (SACT) dataset

Chloe J Bright,[1]* Sarah Lawton,[1] Stephen Benson,[1] Martine Bomb,[1] David Dodwell,[2] Katherine E Henson,[1] Sean McPhail,[1] Louise Miller,[1] Jem Rashbass,[1] Alice Turnbull[1] and Rebecca Smittenaar[1]

[1]National Cancer Registration and Analysis Service, Public Health England, London, UK and [2]Nuffield Department of Population Health, University of Oxford, Oxford, UK

*Corresponding author. National Cancer Registration and Analysis Service, Public Health England, 2 Rivergate, Temple Quay, Bristol BS1 6EH, UK. E-mail: chloe.bright@phe.gov.uk

Data resource basics

Scope

The Systemic Anti-Cancer Therapy (SACT) database is a population-based resource of SACT activity reported routinely by National Health Service (NHS) trusts in England. Data are collected on the SACT treatments of adult and paediatric patients, delivered in secondary and tertiary settings with the intention of improving survival, delaying further cancer progression or development and improving disease-free or progression-free survival. As a guide, the types of treatment NHS trusts are required to submit are listed in Box 1. Treatments such as bisphosphonates and steroids should only be submitted to SACT when they are given with disease-modifying intent (e.g. to control cancer by killing cancer cells and preventing, reducing or delaying cancer growth, development and metastasis), although there is no check on such submissions and some trusts may submit all supportive medications without qualification. It is not within the scope of the SACT database to collect treatments when prescribed as supportive therapies (e.g. to improve symptom control, quality of life or to help manage the side effects of SACT). If a treatment is submitted which would only ever be used with supportive intent, we advise that the treatment is excluded from any analysis. If a treatment is submitted which can be used with supportive or disease-modifying intent, we would assume it is being used in line with the treatment intent (Table 1, field: Drug treatment intent) of the associated regimen. However, we advise analysts that there may be a data quality issue, and recommend that exclusion rules are developed on a case-by-case basis by consideration of the specific treatment and disease indication.

The SACT database is collected and curated by the National Cancer Registration and Analysis Service (NCRAS) at Public Health England (PHE), which started receiving data as part of a phased implementation from April 2012. From April 2014 it became mandatory for all NHS trusts providing SACT to submit data; however, this was

---

Box 1. Examples of systemic anti-cancer therapies that are required and allowed to be submitted to the SACT database. (This is not an exhaustive list)

Systemic anti-cancer therapy	Supportive therapy
• Standard chemotherapy	• Steroids
• Oral chemotherapy	• Bisphosphates
• Immunotherapy	• Antibiotics[a]
• Targeted biological therapies	• Antiemetics[a]
• Endocrine therapies	
• Chimeric antigen receptor T-cell therapy	
• Transcatheter arterial chemoembolization	

[a]These are not required submissions, but are allowed for ease of the submitting trust.

Table 1. Data field descriptions and data quality for all fields collected for use in SACT database

Data field	Description	Data quality issues
Patient table		
NHS number	The primary identifier of a person, is a unique identifier for a patient within the NHS in England and Wales	A small number of records have default codes (0000000000 or 1111111111)
NHS number status indicator code	The verification status of the NHS number provided	No known issues
Date of birth	The date on which a person was born or is officially deemed to have been born	A small volume of records have 01-JAN-1900 as a default date for missing values
Gender code	Person's gender as self-declared (or inferred by observation for those unable to declare their gender)	A small number of gender values are unknown <1%
Ethnicity[a]	The 16 + 1 ethnic data categories defined in the 2001 census, consistent with the national mandatory standard for the collection and analysis of ethnicity	Can be clinician- or patient-reported, which may lead to discrepancies. Some trusts submit additional ethnicity subgroups which are not part of the standard 16 + 1 categories and therefore cannot be reliably interpreted. Recommendation: use ethnicity from other sources (cancer registry data and HES)
Tumour table		
Postcode	The code allocated by the Post Office to identify a group of postal delivery points. Postcodes may also be used to define a geographical area	No known issues
General medical practice code	This is the code of the GP Practice that the patient is registered with	Around 12% of records have either a missing or an unknown code, with a default code of V81999, V81997 or V81998. This has improved in more recent years
Consultant code	Code of consultant who initiated SACT programme	Around 8% of records have a missing GMC[b] code
Care professional main speciality code	Specialty code of consultant who initiated SACT programme	Around 7% of records have either a missing or invalid main specialty code
Organization code	Organization code of the organization acting as a health care provider	A very small number of records have an invalid trust code, e.g. 000, xxx, 101
Primary diagnosis	Primary diagnosis (ICD) at the start of the Systemic Anti-Cancer Therapy	Information in this field is sometimes implausible. Recommendation: use ICD site code from cancer registry data
Morphology	Morphology at time of decision to treat using the International Classification of Disease for Oncology version 3 (ICDO-3)	Some trusts report general morphology codes such as '8000/3' (malignant neoplasm) or '8010/3' (carcinoma, NOS), which may limit the ability to do accurate analysis when looking for particular cancer subtypes. This field has not been verified against data held in the cancer registry data. Recommendation: use morphology from cancer registry data
TNM stage grouping (final pretreatment)	Record the overall clinical TNM stage grouping of the tumour, derived from each T, N and M component before treatment. The overall pretreatment TNM stage grouping indicates the tumour stage at the time the treatment plan was devised	Recording of this field is unreliable due to trusts reporting non-standard staging classifications, e.g. XX1, 100, 310. Recommendation: use stage from cancer registry data when analysing primary cancers
Regimen table		
SACT programme number[a]	Programmes of chemotherapy are numbered according to their chronological order of commencement in the patient's disease management	Some inconsistencies in this field, e.g. some sequential numbers missed. Recommendation: advise caution when using this field
Anti-cancer regimen number[a]	Regimens are numbered according to their chronological order of commencement in the patient's treatment programme	Some inconsistencies in this field, e.g. some sequential numbers missed. Recommendation: advise caution when using this field

(continued)

Table 1. Continued

Data field	Description	Data quality issues
Drug treatment intent	Intent of SACT regimen	In some cases individual clinicians might have used slightly different definitions of curative and palliative treatment
Regimen analysis grouping	SACT regimen group	A small proportion (<1%) is classified as 'NOT MATCHED', e.g. a new regimen that has not been mapped to a regimen analysis grouping yet
Regimen grouping (benchmark reports)	SACT regimen group based on benchmark reports	A small proportion (<1%) is classified as 'NOT MATCHED', e.g. a new regimen that has not been mapped to a regimen benchmark grouping yet
Patient's height (metres)	Height in metres at start of SACT regimen	Some trusts have submitted height using centimetres as there is no upper limit validation. In some instances weight has been wrongly submitted into this field. Oral agents will often not require dosing by height, so may not be recorded
Patient's weight (kilograms)	Weight in kg at start of SACT regimen	Some trusts have submitted weight using grams or stones as there is no upper limit validation (<1%). In some instances height has been wrongly submitted into this field. Oral agents will often not require dosing by weight, so may not be recorded
Performance status	Performance status at start of SACT regimen, adult or young person	Around 6% have invalid codes, e.g. -1. Valid codes for adults = 0-4. Valid codes for young person (<16 years) = 00-11
Comorbidity adjustment indicator	Whether or not patient's overall physical state (other diseases and conditions) was a significant factor in deciding on regimen, or in varying the dose or treatment interval from the start of treatment	Clinicians often interpret this as whether or not the patient has a comorbidity, rather than whether or not the patient's comorbidity influenced their prescribing decision
Decision to treat date (drug regimen)	This is the date on which the consultation between the patient and the clinician took place and a planned cancer treatment was agreed	No known issues
Start date (drug regimen)	This is the first administration date of the first cycle of a regimen	A small volume of records have 01-JAN-1900 as a default date for missing values
Clinical trial indicator	For the SACT programme number, this identifies if a patient's chemotherapy treatment is within a clinical trial	This field is known to miss recognized trial regimens, thus clinical trials are under-reported. Recommendation: advise caution when using this field
Chemo-radiation indicator[a]	This field identifies regimens which are given as part of a combined treatment with radiation	This field is known to miss known regimens including radiotherapy, thus chemo-radiation is under-reported. Recommendation: advise caution when using this field. Use SACT in combination with the radiotherapy dataset to determine chemo-radiation treatment
Cycle table		
Cycle number	Cycles numbered sequentially within each regimen	Some trusts may only report the first cycle of a treatment for a number of reasons, including technical issues with accurate cycle reporting. It may also reflect that patients stopped treatment early due to toxicity, adverse events or patient choice
Start date (cycle)	Date of first drug administration in each cycle	Disproportionately high number of patients receiving first cycle only. This has improved and there were only 22% of records with only one cycle in 2017. Some trusts have submitted '01-JAN-1900'

(continued)

Table 1. Continued

Data field	Description	Data quality issues
Patient's height (metres)	Height in metres at start of SACT cycle	Some trusts have submitted height using centimetres as there is no upper limit validation. In some instances weight has been wrongly submitted into this field. Oral agents will often not require dosing by height, so may not be recorded
Patient's weight (kilograms)	Weight in kg at start of SACT cycle	Some trusts have submitted weight using grams or stones as there is no upper limit validation ($<1\%$). In some instances height has been wrongly submitted into this field. Oral agents will often not require dosing by weight, so may not be recorded
Performance status	Performance status at start of cycle, adult or young person	Around 5% have invalid codes, e.g. -1. Valid codes for adults = 0-4. Valid codes for young person (<16 years) = 00-11
Primary procedure (OPCS)[a]	Procurement code for each administration	No known issues
Drug table		
Drug analysis grouping	The name of the systemic anti-cancer therapy drug group given to a patient during an anti-cancer drug regimen. The name is taken from British National Formulary chapter 8	No known issues
Actual dose	Dose in mg or other applicable unit for each administration in an SACT cycle	Reported units include mg, ml, moles, micromoles and mg/m^2 In addition, for some oral agents the reported dose may be total dose per pack, per 'course' (i.e .multiple cycles) or an estimated value
SACT drug route of administration	The prescribed method of delivery for each administration in an SACT cycle	No known issues
SACT administration date	The date on which the anti-cancer drug was administered to a patient, an infusion commenced, or an oral drug initially dispensed to the patient	No known issues
Organization code (provider)	Code of provider for each administration in a SACT cycle	A small number of records have a missing or invalid trust code, $<2\%$
Primary procedure (OPCS)[a]	Delivery code for each administration	No known issues
Outcomes table		
Start date (final treatment)	The date of the start of the final cycle of SACT treatment within a regimen	Some trusts pre-populate this field as the latest cycle of treatment that has been delivered by the date of data extraction. This may be updated in subsequent data submissions Recommendation: advise caution should be used when using this field
Regimen modification indicator – dose reduction	Identifies if a regimen was modified by reducing the dose of any anti-cancer drug administered at any point in the regimen after commencement of the regimen	No known issues
Regimen modification indicator – time delay[a]	Identifies if a regimen was modified by extending the time between administration dates at any point in the regimen after commencement of the regimen	No known issues
Regimen modification indicator – stopped early[a]	Identifies if a regimen was modified by reducing the administration days below the number planned	No known issues
Planned treatment change reason	To record the immediate outcome of the treatment	No known issues

OPCS, OPCS Classification of Interventions and Procedures; GMC, General Medical Council; TNM tumour, node, metastases staging classification.

[a]These fields have been proposed to be removed by NHS Digital in the next revision of the SACT dataset (Systemic Anti-Cancer Therapy Dataset v3.0).

[b]A code uniquely identifying a general medical practitioner.

not achieved by all trusts until July 2014.[1] SACT data ascertainment has been compared with other national databases.[2,3] Data quality is thought to be sufficient for most purposes from 2013, with the caveat that not all trusts were submitting data until July 2014. Between 1 April 2017 and 31 March 2018, the database recorded a total of 1 220 370 cycles for 209 922 patients. Management of the database is clinically led, and it has been designed to understand patterns in SACT prescribing, and treatment outcomes. The dataset collects information at patient and tumour level and is designed to be linked to other data sources to provide a complete picture of the cancer patient pathway.

Purpose: SACT data for use in research and clinical practice

The SACT dataset design aims to: provide data to support and improve clinical decision making; assess inequalities in access to SACT; and support commissioning [e.g. through providing evaluations for the Cancer Drugs Fund (CDF) and assessing adherence to National Institute for Health and Care Excellence (NICE) guidance]. Summarized SACT data are routinely reported back to the NHS trusts, using an online platform available to users with an N3 connection.[4] The reports available on this platform were developed based on feedback regarding what information trusts would find the most useful in supporting clinical decision making and operational delivery of services. Reports are presented at trust, regional and national levels. In addition the data can be filtered and grouped by specialty (e.g. oncology, haematology, paediatrics), patient age and tumour types.

Examples of data included in the regular trust level reports

- Number of patients receiving treatment for each tumour group.
- Number and proportion of regimens reported for each treatment intent and tumour group.
- Distribution of performance status at start of regimen by treatment intent and tumour group.
- Average number of adults treated by day of the week.
- Number of regimens, number expected to have been completed and number with outcome recorded, stratified by intent of treatment.
- Outcome summary for completed regimens by intent of treatment.
- Names of drugs administered by tumour group.

NCRAS has been commissioned by NHS England (NHSE) to provide data and analysis for the evaluation of new CDF drugs in England.[5] NICE recommends drugs to enter the CDF when there is clinical uncertainty as to
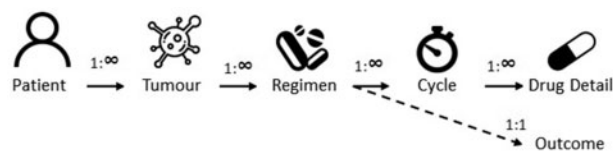

Figure 1. Structure of the tables within the SACT database.

whether the drug should go into routine commissioning. In many circumstances uncertainty will be addressed through a randomized controlled trial (RCT), but SACT and broader cancer data provide a secondary source of information on factors including characteristics of the patient cohort receiving the drug, their overall survival and treatment duration. This information acts to confirm the generalizability of findings from RCTs, by determining if efficacy in clinical trials translates into effectiveness in routine care. In some cases, SACT is used as the primary source for evaluation where the NICE committee believe routine data are sufficient to answer areas of clinical uncertainty in the absence or with support of an RCT. As of October 2018, there were 18 drugs in the Cancer Drugs Fund where SACT and broader NCRAS data will contribute to their evaluation for routine commissioning.[6] As the SACT database does not collect source of funding for systemic therapies, NHSE initially provides additional data to identify the CDF patients.

Structure

The SACT data are stored as six data tables: Patient, Tumour, Regimen, Cycle, Drug detail and Outcome (see Table 1). The SACT data tables have a one-to-many relationship as shown in Figure 1; i.e. a record in the patient table may link to one or many records in the tumour table. The exception is the Outcome table for which a one-to-one relationship with the Regimen table exists; i.e. each completed record in the Regimen table will have a maximum of one record in the Outcome table. Each SACT data table contains a primary key that uniquely identifies records within that table. In addition, each table (except the patient table) contains 'foreign keys' (rows in tables that uniquely define rows in other tables) which are used to link records between tables. For example, the Drug detail table will contain foreign keys to identify which patient, tumour, regimen and cycle the record corresponds to. The internal linkage of the database means that new data on prescribed SACT will automatically be linked to a patient.

Linkage to other data

The SACT database can be linked to other data held by NCRAS, for example the cancer registration data, Hospital Episodes Statistics (HES),[7] the Radiotherapy

Dataset,[8] and the Routes to Diagnosis dataset[9] (see Box 2 for full list). Detailed information on these datasets has been published previously.[10,11] Linkage of these datasets allows the whole patient pathway to be assessed.

Box 2. Datasets and tables available to link through the National Cancer Registration and Analysis Service correct at 17 December 2018[a]

Datasets/tables held by NCRAS
- Cancer registration (including patient, tumour, treatment and income domain table[b])
- Routes to diagnosis table[b]
- Radiotherapy dataset
- Systemic Anti-Cancer Therapy Dataset

Datasets/tables accessible through NCRAS
- Hospital Episode Statistics (including admitted care, out-patient and accident & emergency)
- Diagnostic Imaging Data Set
- National Cancer Waiting Times Monitoring Data
- National Cancer Patient Experience Survey (CPES)
- National Cancer Diagnosis Audit
- The Quality of Life of Cancer Survivors in England
- The Quality of Life of Colorectal Cancer Survivors in England
- LUCADA (Lung Cancer Data Audit (2005-2013))

---

[a]Updated regularly and available at [https://www.gov.uk/government/publications/accessing-public-health-england-data#history].
[b]These are derived data items.

Ascertainment of SACT

Table 2 presents the number of patients treated with a SACT treatment by primary diagnosis, who appear in the SACT dataset compared with the HES-admitted patient care and outpatient care datasets. Overall the ascertainment in SACT is higher compared with HES, especially in adults. In childhood, teenage and young adult (CTYA) patients, HES has higher ascertainment for some sites, particularly in brain/central nervous system (CNS), urology and sarcoma sites. In comparison with adult cancers, CTYA tumours have different approaches to classifying diseases and less extensive roll-out of electronic prescribing systems and associated regimen mapping procedures, as well as complex multicentre models for delivering care. These all contribute to poorer ascertainment of SACT in these cancers. HES inpatient and outpatient data were used in these analyses but, due to the incomplete recording of outpatient procedures, they are likely to underestimate any treatments delivered in an outpatient setting. Only site-specific cancers were included due to a high number of secondary malignant neoplasms captured in HES (C77 – C79) that fell into an 'other' category. This may be because hospital coding teams are attempting to code the metastases as the site of treatment, as opposed to SACT where all treatments are linked to the primary diagnosis at the start of a patient's treatment.

Table 2. Number of patients captured in the Systemic Anti-Cancer Therapies (SACT) dataset vs systemic therapies[a] in HES between 1 March 2017 and 28 February 2018

Cancer site	ICD 10 code	Adults >=25 years		CTYA <25 years	
		HES	SACT	HES	SACT
Brain/CNS	C47, C69-C72	1572	3416	504	447
Breast	C50	26990	33262	30	30
Gynaecological	C51-C58	9350	12035	56	58
Head and neck	C00-C14, C30-C32	3195	3616	35	19
Leukaemia	C91-C95, C962, C964, C968	6473	12051	1709	1721
Lower GI	C18-C21	16371	22852	35	31
Lung	C33-C34, C37-C39, C45	16132	19571	33	18
Lymphoma	C81-C86, C913-C914, C919	14936	16085	701	716
Myeloma	C90	7898	11033	<6[c]	–
Sarcoma	C40-C31, C46, C49	722	1369	406	443
Skin	C43-C44	1682	3678	6	13
Upper GI	C15-C17, C22-C25	11899	14670	52	44
Urology	C60-C68	12054	21882	311	279
Total		129274	175520	3877[b]	3819

CTYA, childhood, teenage and young adult; CNS, central nervous system; GI, gastrointestinal.
[a]Systemic therapy in HES defined using OPCS4 codes: X71, X72, X73, X748, X749, X352, X373, X384.
[b]Total excluding Myeloma- due to small numbers.
[c]non-zero number under 6.

Data collected

Data submission

SACT data are recorded on hospital electronic prescribing systems. Patient details and the treatment prescribed are entered during the course of care by clinicians, nurses, pharmacists and other health care providers. There are several electronic prescribing software providers, but all produce a SACT extract in a standard format. Trusts make a data extract monthly according to their submission schedule. The few trusts without electronic prescribing use Patient Administration Systems or a manual system to produce their extracts. SACT data extracts are manually uploaded to a secure portal maintained by PHE (there is no direct interface between the portal and electronic prescribing system). Each trust has at least one registered uploader who is responsible for uploading the extract.

If there are any validation errors on the extract, for example missing data in key fields, the uploader must correct these and re-upload the data. Support with errors and the upload process is provided by the SACT helpdesk. A summary report of total administrations, new cycles and regimens is provided on the portal to enable trust users to check the data before submitting the extract.

Clinical and patient data

Table 1 provides complete information on the 43 data fields and their description, as well as any data quality issues associated with these. Table 3 provides information on the completeness of each data field by cancer type for the period 1 April 2017 to 31 March 2018. Data completeness has improved over the past years and, as such, cohorts including historical data will not have such complete data.

The topography of the tumour is coded using the International Classification of Disease 10th Revision,[12] and the histology is coded using the International Classification of Disease Oncology 3rd revision.[13] These codes refer to the primary diagnosis of the tumour and the description of the tumour at the start of SACT treatment.

Including patient's postcode as a field means that information can be mapped to other health care geographies, to support service planning within Clinical Commissioning Groups or Cancer Alliances. In addition, the patient's postcode allows the deprivation score to be determined using quintiles of income domain at the Lower Super Output Area, enabling the investigation of inequalities in prescribing of SACT.

Data collection updates

A review of the dataset is currently being undertaken. Version 3.0 will be an extension to the standard, introducing new data fields, correcting existing data fields and removing some data fields to reduce the burden of data collection wherever possible. This is to update the NHS Digital Information Standard ISB 1533, to standardize the recording of systemic anti-cancer therapy treatments and outcomes through electronic systems. A phased implementation of the revised data standard will start from September 2019 with full implementation by December 2019.

Data resource use

The SACT dataset has been used to investigate 30-day mortality following SACT for lung and breast cancer patients.[14] This research provided real-world evidence on the 30-day mortality of patients receiving SACT, as well as demonstrating variation between trusts. Additionally, NCRAS have produced a methodology and process for routine feeds to trusts to highlight patients who have died within 30 days of receiving SACT, to potentially support morbidity and mortality meetings. Reviews of individual deaths have been highlighted as a way to increase the quality of care.[15] NCRAS plan to develop additional outcome metrics to feedback to trusts to support the delivery of care.

Henson *et al.*[16] used the SACT dataset when investigating sociodemographic variation in the use of chemotherapy and radiotherapy in patients with stage IV lung, oesophageal, stomach or pancreatic cancer. This work has been further developed into an online tool[17] and workbooks presenting the data.[18,19] Jones *et al.*[20] used the SACT dataset when investigating 30-day mortality after chemotherapy for people with small-cell lung cancer. Annual reports are published for the National Cancer Audits[21–23]; these report on SACT activity with the overall aim to review the quality of cancer care and identify areas for improvement. A list of all published reports can be found on the SACT dataset website.[24] Ongoing research projects are investigating whether age is a barrier to receiving SACT,[25] and examining the relationship between SACT prescription in a palliative setting and end-of-life care.[26] Health Data Insight CiC (HDI), in partnership with AstraZeneca and IQVIA, have developed a dataset (the Simulacrum) containing artificial patient-like cancer data, which imitates some of the data held by NCRAS, including the SACT dataset. More information is available on the HDI website.[27] The SACT and NCRAS websites contains details on the methodologies used in our various outputs.[2,28]

Strengths and weaknesses

Strengths

Coverage and clinical detail
The key strength of the SACT database is the detailed clinical information collected with whole population coverage

Table 3. Data field completeness by cancer type for the period 1 April 2017 to 31 March 2018, the ICD-10 codes for the groups are consistent with Table 2

Diagnostic group	Number of patients	Number of tumour records	Number of regimens	Number of cycles	Number of drug records	Number of outcome records	% NHS number	% Date of birth	% Current gender	% Ethnicity	% Patient postcode	% GP practice code	% GMC code[b]	% Consultant Specialty[b]	% Primary diagnosis
Brain/CNS	3519	3563	5548	14 588	26 387	5000	100%	100%	100%	88%	100%	96%	99%	100%	100%
Breast	41 165	42 761	77 142	297 014	606 089	70 209	100%	100%	100%	87%	100%	95%	99%	100%	100%
Gynaecological	12 620	13 117	18 645	64 801	186 589	17 088	100%	100%	100%	89%	100%	96%	99%	100%	100%
Head and neck	3570	3601	4907	14 264	46 877	4476	100%	100%	100%	84%	100%	94%	98%	100%	100%
Leukaemia	10 756	11 098	16 217	50 977	119 424	15 164	100%	100%	100%	91%	100%	95%	97%	100%	100%
Lower GI	23 062	23 961	33 835	141 250	441 456	30 779	100%	100%	100%	85%	100%	94%	98%	100%	100%
Lung	19 811	20 472	27 872	90 065	224 699	25 343	100%	100%	100%	85%	100%	95%	99%	100%	100%
Lymphoma	17 569	18 512	27 078	80 191	328 896	25 068	100%	100%	100%	90%	100%	94%	97%	100%	100%
Miscellaneous[a]	16 917	17 551	21 644	72 011	136 835	19 948	100%	100%	100%	90%	100%	95%	96%	100%	100%
Myeloma	12 970	13 532	24 295	88 118	228 451	22 735	100%	100%	100%	93%	100%	94%	98%	100%	100%
Sarcoma	1385	1418	2036	6852	17 698	1812	100%	100%	100%	91%	100%	98%	99%	100%	100%
Skin	3747	3809	4984	25 269	34 431	4540	100%	100%	100%	89%	100%	93%	99%	100%	100%
Upper GI	14 768	15 350	20 488	65 713	229 747	18 472	100%	100%	100%	85%	100%	95%	98%	100%	100%
Urology	26 085	26 773	35 353	124 912	224 722	31 993	100%	100%	100%	86%	100%	95%	97%	100%	100%
Paediatrics (under 16)	2801	2927	7348	16 259	80 219	5472	100%	100%	100%	92%	100%	96%	99%	100%	100%
TYA (16–23)	1599	1677	3082	8267	36 404	2669	100%	100%	100%	88%	100%	98%	99%	100%	99%

Diagnostic group	% Morphology	% Stage of disease at start of programme	% Programme number	% Regimen number	% Treatment intent	% Regimen name	% Height at start of regimen	% Weight at start of regimen	% Performance status at start of regimen	% Comorbidity adjustment	% Date of decision to treat	% Start date of regimen	% Clinical trial	% Chemo radiation	% Number of cycles planned	% Cycle number
Brain/CNS	67%	30%	97%	89%	99%	100%	95%	92%	81%	75%	90%	100%	98%	95%	89%	100%
Breast	64%	70%	96%	90%	89%	100%	86%	86%	82%	79%	94%	100%	97%	95%	90%	100%
Gynaecological	65%	68%	94%	89%	97%	100%	95%	95%	83%	77%	96%	100%	98%	95%	93%	100%
Head and neck	63%	74%	92%	91%	98%	100%	97%	96%	81%	74%	91%	100%	94%	90%	95%	100%
Leukaemia	56%	19%	97%	93%	93%	100%	65%	67%	67%	83%	94%	100%	97%	95%	90%	100%
Lower GI	59%	73%	94%	91%	97%	100%	95%	95%	84%	81%	94%	100%	98%	93%	92%	100%
Lung	75%	82%	96%	93%	96%	100%	90%	91%	86%	82%	95%	100%	98%	94%	91%	100%
Lymphoma	58%	37%	96%	93%	93%	100%	90%	90%	74%	83%	94%	100%	97%	95%	93%	100%
Miscellaneous[a]	48%	15%	94%	92%	88%	100%	49%	53%	64%	82%	94%	100%	97%	93%	89%	100%
Myeloma	62%	29%	96%	90%	97%	100%	83%	84%	68%	85%	95%	100%	98%	95%	89%	100%
Sarcoma	55%	38%	92%	85%	98%	100%	88%	87%	83%	62%	92%	100%	91%	86%	90%	100%
Skin	54%	55%	95%	89%	96%	100%	77%	86%	85%	81%	93%	100%	98%	94%	88%	100%
Upper GI	55%	68%	94%	89%	94%	100%	90%	91%	83%	77%	94%	100%	97%	93%	90%	100%
Urology	59%	62%	94%	92%	98%	100%	64%	65%	75%	78%	94%	100%	96%	94%	88%	100%
Paediatrics (under 16)	67%	8%	99%	77%	94%	100%	31%	92%	10%	67%	99%	100%	94%	78%	88%	100%
TYA (16–23)	60%	31%	93%	84%	96%	100%	75%	93%	66%	60%	90%	100%	90%	82%	89%	100%

Diagnostic Group	% Weight at start of cycle	% Performance status at start of cycle	% OPCS procurement code	% Drug name	% Actual dose per administration	% Administration route	% Administration date	% OPCS delivery code	% Organization code of drug provider	% Date of final treatment	% Regimen modification (dose reduction)	% Regimen modification (time delay)	% Regimen modification (stopped early)	% Regimen outcome summary
Brain/CNS	84%	70%	88%	100%	94%	98%	100%	87%	100%	34%	86%	70%	83%	11%
Breast	80%	80%	63%	100%	96%	98%	100%	69%	99%	39%	82%	75%	81%	12%
Gynaecological	92%	80%	82%	100%	97%	99%	100%	84%	100%	39%	85%	75%	82%	16%
Head and neck	95%	81%	75%	100%	94%	99%	100%	80%	99%	46%	80%	69%	82%	21%
Leukaemia	62%	69%	70%	100%	97%	99%	100%	67%	99%	40%	85%	75%	83%	7%
Lower GI	94%	83%	84%	100%	95%	98%	100%	85%	99%	34%	83%	76%	81%	15%
Lung	87%	82%	71%	100%	95%	98%	100%	81%	99%	36%	83%	74%	81%	14%
Lymphoma	89%	73%	75%	100%	97%	99%	100%	78%	99%	43%	83%	73%	81%	10%
Miscellaneous[a]	56%	67%	71%	100%	93%	99%	100%	75%	99%	41%	80%	70%	78%	6%
Myeloma	77%	72%	52%	100%	96%	98%	100%	63%	99%	37%	85%	75%	84%	10%
Sarcoma	87%	85%	75%	100%	98%	100%	100%	70%	100%	40%	82%	62%	86%	10%
Skin	86%	82%	55%	100%	96%	98%	100%	59%	100%	26%	85%	78%	84%	11%
Upper GI	88%	81%	80%	100%	95%	98%	100%	85%	99%	35%	81%	73%	79%	14%
Urology	67%	75%	56%	100%	96%	98%	100%	68%	99%	37%	76%	69%	75%	10%
Paediatrics (under 16)	93%	8%	71%	100%	96%	99%	100%	78%	100%	55%	68%	64%	67%	4%
TYA (16-23)	89%	67%	75%	100%	98%	100%	100%	78%	100%	50%	79%	55%	82%	11%

CNS, central nervous system; GI, gastrointestinal; TYA, teenage or young person.

[a]Any valid ICD 10 not listed in specific group.

[b]Code uniquely identifying a general medical practitioner.

for England. The completeness and population-based design of the data source minimizes selection bias and increases the external validity of findings from studies using these data.[29] Following PubMed searches and through conversations with international colleagues working in the field of cancer data, we have been unable to find a data source which is comparable to the SACT database in terms of its breadth or depth. NHS Scotland also collect data on SACT prescribed to patients, which can be linked to their cancer registry data, and includes laboratory test results. However, the population of Scotland is 5.5 million compared with 55 million in England, and the population coverage of the dataset is approximately 50%; therefore, there are considerably fewer patients in the dataset. Additionally, these data are not yet being made available to external researchers. Other similar international data sources are restricted by insurance status, only cover specific provinces or states or are not dedicated resources for the collection of SACT.[30–37]

Routine population-based databases such as SACT can be used to ascertain whether findings from RCTs generalize into 'real-world' settings. Participant groups recruited to RCTs are commonly small and unrepresentative of the broader population of cancer patients.[38–40] There may also be differences in care between patients enrolled in an RCT compared with patients in routine NHS practice. This is particularly important when the RCT-reported efficacy of a new cancer treatment is marginal, and the real-world outcomes for patients may deviate substantially from this. Population level data also facilitate reporting on serious adverse events from these treatments, which is critical given that there is concern that RCTs underestimate treatment toxicity.[41,42] Routine population-level databases can also be used to complete research into rare cancers where RCTs may not be practical.

Granularity of data collection
Data should be entered into the SACT portal each and every time SACT is prescribed, as well as the outcomes associated with treatment. Such granularity of data collection enables in-depth exploration of all the drugs prescribed to patients throughout the course of their treatment.

Data linkage
The ability to link to a variety of routine care databases collected by the NHS (see Box 2) greatly increases the granularity in which analyses can be conducted on different patient characteristics, tumour types and care pathways. Correspondingly, this enables a vast number of research questions to be addressed, such as inequalities in access to treatments. The use of primary and foreign keys within the data tables allows all records of SACT for a cycle, regimen, tumour and patient to be linked over time. This longitudinal linkage allows long-

term follow-up of patients to death; in addition to variations in prescribing practice over time.

Weaknesses

Many of the data issues are a result of this being a real-world database, collected routinely at the point of care. NCRAS works with trusts to support them in uploading the data, as well as supporting their transition to electronic prescribing systems. Over time, the number of data validations which are performed on the data before submission have increased.[43] In the current system, records that fail the SACT portal validation process appear as critical errors and must be corrected before the file can be submitted. Data quality is also addressed when SACT data is linked to the cancer registry dataset. If any issues are identified, the SACT team have data liaison officers who support trusts to improve the quality of the data they submit. The team also has a newsletter, a helpdesk and a 'knowledge hub' to increase engagement and knowledge sharing regarding the dataset. Nevertheless, the source data are entered in real-time through electronic prescribing systems by clinicians during busy clinics, often when the patient is present. Full and accurate information may not be readily available, and there may not be time available to search other clinical records for this information. Substantial resource, therefore, has to be invested by NCRAS in processing the data before using it for analysis.

Ascertainment

As shown in the comparison with HES data, ascertainment is poorer for CTYA cancers. Ascertainment of SACT can be heavily linked to the trusts' use of electronic prescribing systems. For CTYA trusts and centres, the initiative to use electronic prescribing was much later than for adult services. We expect to see an improvement in ascertainment of SACT among CTYA due to the increased use of electronic prescribing systems.

For several trust-level operational reasons, ascertainment is also poorer for oral chemotherapy. One example would be that some trusts do not use electronic prescribing systems for oral drugs. Drugs which are delivered 'outside' an oncology environment (e.g. in surgical clinics or in primary care) are often poorly recorded. The most pertinent example is the use endocrine therapies in breast and prostate cancer.

Data quality

Cycle number and programme number are sometimes missing or incorrectly submitted. It is possible to infer cycle number by using the cycle date to sequentially order the SACT prescribed; however, this field was not mandatory

until June 2017. Determining whether a patient has changed programme (typically called 'line of treatment' by clinicians) can be inferred with support from clinicians based upon their knowledge of the clinical pathway. Data quality for important fields such as stage and morphology are also poor (see Table 1). However, this can be supplemented by linking to cancer registry data.

When inputting data, clinicians are able to name the regimens as they wish. Each trust will have a pharmacist or data uploader who is responsible for mapping the clinician-generated regimens to a list of nationally approved regimens. Whilst this only needs to be done once (the same regimen uploaded again will be automatically mapped), it is a resource-intensive process.

Before electronic prescribing, there was a disproportionately high number of patients receiving 'first cycle only' chemotherapy. It is problematic to infer from these data which patients received only a single cycle of SACT, and then stopped treatment, or whether this is a data quality issue and the patient received multiple cycles of SACT which were not recorded.

A comprehensive list of data quality issues has been provided in Table 1. As indicated, some fields can be supplemented with information from the NCRAS tables, thus overcoming these issues.

Data resource access

Section 251 of the NHS Act 2006 grants PHE permission to collect information on cancer patients for health improvement and service provision purposes, without the need to seek consent. These provisions are reviewed annually by the Confidentiality Advisory Group of the Health Research Authority.[44]

Applications to access potentially identifiable SACT or other NCRAS data can be made to the Office for Data Release (ODR) at PHE [odr@phe.gov.uk]. There are three stages to the ODR approval process: (i) application; (ii) assessment and approval; and (iii) access. It is strongly recommended that prospective applicants discuss their proposed project with the ODR before submission of a full application. Further details regarding ODR application processes are available online.[45] The NCRAS data dictionary [https://www.gov.uk/government/publications/accessing-public-health-england-data] details data items which are available for access request, and highlights which data items are potentially identifiable and therefore not available to researchers. If linkage to other NCRAS datasets is requested, linkage will be undertaken by an NCRAS analyst before release of data.

There are no restrictions on who can apply to the ODR; however, applicants are required to demonstrate that they

are compliant with UK data protection laws including the Common Law Duty of Confidentiality,[46] the General Data Protection Regulation (EU) 2016/679,[47] and the seven Caldicott Principles.[48] Furthermore, applicants must demonstrate that their purpose for using the data is consistent with the aims of health improvement and service provision, in line with the permissions to collect the data under Section 251 of the NHS Act 2006. Depending on whether the work is deemed to be research or service evaluation, applicants will also have to demonstrate approval from a research ethics committee associated with their institution.

All requests to access these data will be charged on a project basis, to reflect the amount of work required to facilitate access to requested data. Charges will be kept as low as possible and aim to cover costs rather than produce profit. Pre-application advice is offered free of charge and aims to provide expert advice to help prospective applicants. Further details regarding costs are available online.[49]

The SACT helpdesk is available at [SACT@phe.gov.uk] to help support trusts with data submission, and to receive feedback from trusts on the routine reporting. In addition, data dictionaries and user guidance are available online.[1,24]

Conclusion

The SACT dataset is a unique resource in terms of the breadth and depth of information on systemic anti-cancer drugs prescribed to cancer patients. The key strengths are the population-based nature, high level of clinical information beyond standard cancer registration data and the ability to link to other NCRAS datasets providing the opportunity to have a complete picture of the patient pathway. Considerable resource has been invested to increase data quality; however, issues relating to ascertainment and completeness of certain fields still persist. Nevertheless, a wealth of research questions can be addressed using this dataset, which will vastly improve our knowledge of SACT prescribing practice and, variation in this, as well as the associated treatment outcomes.

Profile in a nutshell

- The Systemic Anti-Cancer Therapy (SACT) dataset is an ongoing mandated collection of all SACT prescribed to cancer patients by secondary and tertiary care providers in the English National Health Service (NHS).
- Data are collected on patient demographics, the consultant and NHS trust where the patient is being treated, tumour characteristics, the clinical status of the patient and the SACT the patient is receiving.
- SACT data can be linked to several datasets held by the National Cancer Registration and Analysis Service (NCRAS) and Hospital Episode Statistics (HES) to provide a more complete picture of the patient pathway.
- SACT data are being used for the appraisal of the drugs funded through the Cancer Drugs Fund and for a variety of observational research purposes, including investigating 30-day mortality and inequalities in access to treatment.
- NCRAS routinely support provider trusts with their data submissions, focusing on improving data quality and completeness. The NHS England Medicines Optimisation Commissioning for Quality and Innovation framework has supported NHS trusts in their efforts to improve data quality for a number of key items.
- Understanding the clinical context to SACT prescribing is key in correctly interpreting and maximizing the value of these data.

Glossary

CDF: Cancer Drugs Fund

CTYA: Childhood, teenage and young adult cancer. Defined as a cancer diagnosed when aged <25 years

Electronic prescribing system: Sends electronic prescriptions form GP surgeries directly to pharmacies

Foreign key: A field in one table that uniquely identifies a row in another table

HES: Hospital Episodes Statistics dataset. This dataset includes patient and clinical information on hospitaladmissions and attendances

ICD-10: The International Classification for Disease version 10. This is the international standard for classifying diseases and is published by the World Health organization.

ICDO-3: The International Classification for Disease Oncology version 3. This is the international standard for classifying diseases in oncology.

PHE: Public Health England

Primary key: A field which contains a unique identifier

NCRAS: National Cancer Registration and Analysis Service. This service is part of Public Health England andis responsible for all cancer registrations in England

NHS: National Health Service

RCT: Randomized controlled trial

SACT: Systemic Anti-Cancer Therapy, includes chemotherapy, immunotherapy and biological therapies

Trust: NHS foundation trusts and acute trusts ensure hospitals provide high-quality health care. The majority of hospitals in England are now managed by NHS foundation trusts. NHS trusts are independent legal entities and have unique governance arrangements

References

1. Systemic Anti-Cancer Therapy team. *Systemic Anti-Cancer Therapy Dataset (SACT) Frequently Asked Questions*. 2018. http://www.chemodataset.nhs.uk/frequently_asked_questions/ (31 August 2018, date last accessed).
2. National Cancer Registration and Analysis Service. *Chemotherapy, Radiotherapy and Surgical Tumour Resections in England*. http://www.ncin.org.uk/cancer_type_and_topic_specific_work/topic_specific_work/main_cancer_treatments (10 August 2018, date last accessed).
3. National Cancer Registration and Analysis Service. *Matching SACT to Cancer Waiting Times Data*. 2014. http://www.ncin.org.uk/publications/data_briefings/sact_cwt (10 April 2019, date last accessed).
4. National Cancer Registration and Analysis Service. CancerStats. https://nww.cancerstats.nhs.uk/users/sign_in (20 June 2019, date last accessed).
5. NHS England. *Cancer Drugs Fund: NHS England*. https://www.england.nhs.uk/cancer/cdf/ (31 August 2018, date last accessed).
6. NHS England. *National Cancer Drugs Fund List*. 2018. https://www.england.nhs.uk/wp-content/uploads/2018/10/national-cdf-list-v1.105.pdf (31 August 2018, date last accessed).
7. NHS Digital. 2019. *Hospital Episode Statistics*. https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics (20 June 2019, date last accessed).
8. National Cancer Registration and Analysis Service. *Radiotherapy Dataset (RTDS)*. http://www.ncin.org.uk/collecting_and_using_data/rtds (31 August 2018, date last accessed).
9. 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.
10. Public Health England. *PHE Cancer Data Sets, Linkage and Availability*. 2016.
11. Henson KE, Elliss-Brookes L, Coupland VH *et al*. Data Resource Profile: National Cancer Registration Dataset in England. *Int J Epidemiol* 2020;**49**:16–16h.
12. World Health Organization. *International Classification of Diseases - 10th Revision*. Geneva: WHO, 1990.
13. World Health Organization. *International Classification of Diseases for Oncology*. 3rd edn, 1st revision. Geneva: WHO, 2013.
14. Wallington M, Saxon EB, Bomb M *et al*. 30-day mortality after systemic anticancer treatment for breast and lung cancer in England: a population-based, observational study. *Lancet Oncol* 2016;**17**:1203–16.
15. Hogan H, Zipfel R, Neuburger J, Hutchings A, Darzi A, Black N. Avoidability of hospital deaths and association with hospital-wide mortality ratios: retrospective case record review and regression analysis. *BMJ* 2015;**351**:h3239.
16. 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.
17. National Cancer Registration and Analysis Service and Cancer Research UK. *Chemotherapy, Radiotherapy and Surgical Tumour Resections in England*. London: Public Health England, NCRAS, 2018.
18. Chemotherapy, Radiotherapy and Tumour Resection by Tumour & Patient Characteristics in England, 2013–2015, 2018. http://www.ncin.org.uk/cancer_type_and_topic_specific_work/topic_specific_work/main_cancer_treatments (August 10, 2018, date last accessed).
19. Cancer Alliance. *Chemotherapy, Radiotherapy and Tumour Resection by Cancer Alliance in England, 2013–2015*, 2018. http://www.ncin.org.uk/cancer_type_and_topic_specific_work/topic_specific_work/main_cancer_treatments (August 10, 2018, date last acces sed).
20. Jones GS, McKeever TM, Hubbard RB, Khakwani A, Baldwin DR. Factors influencing treatment selection and 30-day mortality after chemotherapy for people with small-cell lung cancer: an analysis of national audit data. *Eur J Cancer* 2018;**103**:176–83.
21. Royal College of Physicians. *National Lung Cancer Audit Royal College of Physicians*. https://www.rcplondon.ac.uk/projects/national-lung-cancer-audit (4 December 2018, date last accessed).
22. Royal College of Surgeons of England. National Prostate Cancer Audit. https://www.npca.org.uk/ (4 December 2018, date last accessed).
23. Clinical Effectiveness Unit Royal College of Surgeons of England. *National Audit of Breast Cancer in Older Patients*. https://www.nabcop.org.uk/reports/ (4 December 2018, date last accessed).
24. Systemic Anti-Cancer Therapy team. *SACT Data Reports*. http://www.chemodataset.nhs.uk/reports/ (31 August 2018, date last accessed).
25. Wallington M, Battisti N, Ring A *et al*. Is age a barrier to chemotherapy? Rates of treatment in older people with breast, colon or lung cancer in England in 2013–15: a national registry study. In: *The European Society for Medical Oncology Congress, 19-23 October 2018, Munich, Germany*. Lugano, Switzerland: ESMO, 2018.
26. Dunlop C, Wallington M, Charman J *et al*. What proportion of pancreatic cancer patients are prescribed treatment beyond first line palliative chemotherapy? Deriving line of chemotherapy from SACT data. In: *PHE Cancer Services, Data and Outcomes Conference, 20-21 June 2018, Manchester, UK*. London: Public Health England, 2018.
27. Health Data Insight. *The Simulacrum*. 2018. https://simulacrum.healthdatainsight.org.uk/ (10 April 2019, date last accessed).

28. SACT team. *SACT Systemic Anti-Cancer Therapy Dataset*. http://www.chemodataset.nhs.uk/home (31 August 2018, date last accessed).

29. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. *Br J Cancer* 2014;**110**:551–55.

30. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care* 2002;**40**:3–18.

31. National Cancer Institute. *Patterns of Care/Quality of Care Studies*.

32. Scottish Cancer Registry. *Scottish Open Cancer Registration And Tumour Enumeration System (SOCRATES)*. Edinburgh: SCR. https://www.isdscotland.org/Health-Topics/Cancer/Scottish-Cancer-Registry.asp (20 June 2019, date last accessed).

33. CancerCare Manitoba. Manitoba Cancer Registry and Treatment. Winnipeg, MB: CancerCare Manitoba,

34. ISD Scotland. *General Acute Inpatient and Day Case - Scottish Mortality Record* (SMR01). Edinburgh: ISD Sccotland,

35. Ontario CC. *Ontario Cancer Registry*. Toronto, ON: Ontario Cancer Care.

36. Danish Health Data Authority. *The National Patient Register*. Copenhagen: National Centre for Register-based Research.

37. Victoria State Government. *Victorian Admitted Episodes Dataset*. Melbourne, VIC: Victorian Government, 2018.

38. Hutchins LF, Unger JM, Crowley JJ, Coltman CA Jr, Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *N Engl J Med* 1999;**341**:2061–67.

39. Lewis JH, Kilgore ML, Goldman DP *et al*. Participation of patients 65 years of age or older in cancer clinical trials. *J Clin Oncol* 2003;**21**:1383–89.

40. Unger JM, Hershman DL, Albain KS *et al*. Patient income level and cancer clinical trial participation. *J Clin Oncol* 2013;**31**: 536–42.

41. Seruga B, Sterling L, Wang L, Tannock IF. Reporting of serious adverse drug reactions of targeted anticancer agents in pivotal phase III clinical trials. *JCLin Oncol* 2011;**29**:174–85.

42. Niraula S, Seruga B, Ocana A *et al*. The price we pay for progress: a meta-analysis of harms of newly approved anticancer drugs. *J Clin Oncol* 2012;**30**:3012–19.

43. Systemic Anti-Cancer Therapy team. *New Quality and Completeness Checks on Submissions of Systemic Anti-Cancer Therapies (SACT) Data*. London: Public Health England, 2017.

44. NHS Health Research Authority. *Section 251 and the Confidentiality Advisory Group (CAG)*. http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/ (31 August 2018, date last accessed).

45. Public Health England. *The Office for Data Release*. https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-phe-odr-and-accessing-data (20 June 2019, date last accessed).

46. van Leeuwen FE, Klokman WJ, Stovall M *et al*. Roles of radiation dose, chemotherapy, and hormonal factors in breast cancer following Hodgkin's disease. *J Natl Cancer Inst* 2003;**95**:971–80.

47. European Union. *General Data Protection Regulation*, Brussels: European Parliament and Council of the European Union, 2018.

48. Department of Health. *Information: To share or not to share? The Information Governance Review*, London: DoH, 2013.

49. Public Health England. *Office for Data Release: Services and Products Cost Recovery*. 2017. https://www.gov.uk/government/publications/office-for-data-release-services-and-products-cost-recovery.