Medicines & Healthcare products Regulatory Agency

CPRD

Systemic Anti-Cancer Therapy (SACT) Data Dictionary

Version 5.1

Date: 26 March 2021

NIHR | National Institute for Health Research

Documentation Control Sheet

Over time, it may be necessary to issue amendments or clarifications to parts of this document. This form must be updated whenever changes are made.

Version	Summary of Change	Prepared By	Reviewed By
1.0	Initial draft	Helen Strongman	Rachael Williams
2.0	Modified	Rachael Williams	Eleanor Yelland
3.0	Modified	Helen Booth	Eleanor Yelland
4.0	Modified	Sonia Coton	Eleanor Yelland
5.0	Modified	Eleanor Yelland	Hilary Shepherd
5.1	Modified	Hilary Shepherd	

Version 1.0
- First version (20 June 2017) for set 14
- No changes for set 15

Version 2.0
- Updated for set 16

Version 3.0
- Updated for set 17

Version 4.0
- Updated for set 18

Version 5.0
- Updated for set 19

Version 5.1
- Updated for set 21
- Updated branding
- Added DOIs
- Added v3 SACT variables: regoutsum_cur_not_com_plan, regoutsum_non_curat, regoutsum_toxic, regoutsum_cur_com_plan, adjunctive therapy
- Dropped v2 SACT variables: stage_at start, programme_number, regimen_number, chemo_radiation, number_of_cycles_planned, opcs_procurement_code, regimen_modification_time_delay, regimen_modification_stopped_early, regimen_outcome_summary, opcs_delivery_code, date_of_final_treatment, organisation_code_of_provider

DOI
Please cite in any publications using these data:

CPRD GOLD SACT August 2021 - https://doi.org/10.48329/4jwp-4951

CPRD Aurum SACT August 2021 - https://doi.org/10.48329/ngt7-ex36

1. **Demographics and Consultant** - One row per patient

Column description	*Column Name*	*Details*	*Field Type*	*Valid Content*
CPRD patient identifier	e_patid	Unique patient identifier based on CPRD primary care data	ID	Number
CR patient Identifier	e_cr_patid	Unique patient identifier based on NCRAS data. In some cases, the same person may have multiple patient IDs. Patient IDs will be retained even after two patient records are found to be the same person.	ID	Number
Pseudonymised Tumour Identifier	e_cr_id	Tumour Identifier	NUMBER	Max length 6 characters
Consultant code	consultant_gmc_code	Pseudonymised code of the consultant who initiated the SACT treatment	NUMBER	Number
Consultant Specialty Code	consultant_speciality_code	Main specialty code of the consultant who initiated the SACT treatment	NUMBER	Max Length 8 characters, Range 0 - 991

2. Clinical Status

Column description	*Column Name*	*Details*	*Field Type*	*Valid Content*
CPRD patient identifier	e_patid	Unique patient identifier based on CPRD primary care data	ID	Number
CR patient Identifier	e_cr_patid	Unique patient identifier based on NCRAS data. In some cases, the same person may have multiple patient IDs. Patient IDs will be retained even after two patient records are found to be the same person.	ID	Number
Primary Diagnosis	primary_diagnosis	Primary diagnosis at the initiation of SACT (at the time of decision to treat) (ICD-10)	TEXT	Max Length 6 characters
Morphology	morphology_clean	Morphology ICD-O at the initiation of SACT (at the time of decision to treat)	TEXT	Max Length 5 characters

3. Programme and Regimen

Column description	*Column Name*	*Details*	*Field Type*	*Valid Content*
CR patient Identifier	e_cr_patid	Unique patient identifier based on NCRAS data. In some cases, the same person may have multiple patient IDs. Patient IDs will be retained even after two patient records are found to be the same person.	ID	Number
Pseudonymised Tumour Identifier	e_pseudo_merged_tumour_id	Tumour Identifier	NUMBER	Max length 6 characters
Pseudonymised Regimen identifier	e_pseudo_merged_regimen_id	Regimen Identifier (used to link to all subsequent tables)	NUMBER	Max length 6 characters
Drug treatment intent	intent_of_treatment	Intent of SACT regimen	TEXT	01 = Curative - aiming to permanently eradicate disease 02 = Palliative - Aiming to extend life expectancy 03 = Palliative - Aiming to relieve and/or control malignancy related symptoms 04 = Palliative - Aiming to achieve remission 05 = Palliative - Aiming to delay tumour progression 98 = Other 99 = Not Known
Adjunctive therapy	adjunctive_therapy	Adjunctive therapy	TEXT	1= Adjuvant 2= Neoadjuvant 3= Not Applicable (Primary Treatment) 4= Not Known
Regimen Analysis Group	analysis_group	This is a cleaned version of variables mapping regimen information into consistent description groups. A SACT regimen identifies a standard for a combination of drugs (or single drug) given	TEXT	Max Length 59 characters

		in a planned schedule Note: includes ‘Not Chemo’, ‘Not matched’.		
Regimen Benchmark Group	benchmark_group	Benchmark_Group - Maps regimen information into consistent high-level groups (e.g. ‘CHOP R – 21 days’ is mapped to CHOP R or DEGARELIX, BICALUTAMIDE + GOSERELIN and TAMOXIFEN are mapped to HORMONES). Note includes ‘Not Chemo’, ‘Not matched’.	TEXT	Max Length 59 characters
Person Height at start of Regimen	height_at_start_of_regimen	Height in metres at the start of the SACT Regimen	NUMBER	Range -1 to 9.99 Max 2 d.p
Person Weight at start of Regimen	weight_at_start_of_regimen	Weight in kilograms at the start of the SACT Regimen	NUMBER	Range -1 to 999 Max 3 d.p
Performance Status at Start of Regimen	perf_status_start_of_regimen_clean	Performance Status Indicator of Person at start of regimen (adult)	TEXT	0 = Able to carry out all normal activity without restriction 1 = Restricted in physically strenuous activity, but able to walk and do light work 2 = Able to walk and capable of all self care, but unable to carry out any work. Up and about more than 50% of waking hours 3 = Capable of only limited self care, confined to bed or chair more than 50% of waking hours 4 = Completely disabled. Cannot carry on any self care. Totally confined to bed or chair Performance status (adult and young person)- SACT

				version 2 0 = Able to carry out all normal activity without restriction 1 = Restricted in physically strenuous activity, but able to walk and do light work 2 = Able to walk and capable of all self care, but unable to carry out any work. Up and about more than 50% of waking hours 3 = Capable of only limited self care, confined to bed or chair more than 50% of waking hours 4 = Completely disabled. Cannot carry on any self care. Totally confined to bed or chair 00 = 100% - Fully active, normal 01 = 90% - Minor restrictions in physically strenuous activity 02 = 80% - Active, but tires more quickly 03 = 70% - Both greater restriction of, and less time spent in, play activities 04 = 60% - Up and around, but minimal active play; keep busy with quieter activities 05 = 50% - Gets dressed but lies around much of

				the day; no active play; able to participate in all quiet play and activities 06 = 40% - Mostly in bed; participates in quiet activities 07 = 30% - In bed; needs assistance even for quiet play 08 = 20% - Often sleeping; play entirely limited to very passive activities 09 = 10% - No play; does not get out of bed 10 = 5% - Unresponsive 11 = 0% - Dead
Comorbidity Adjustment Indicator	comorbidity_adjustment	Indicator of 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. Y = Yes; N= No	TEXT	Y = Yes, N = No
Decision to Treat Date (Drug Regimen)	date_decision_to_treat	This is the date that the consultation between the patient and the clinician took place and a Planned Cancer Treatment was agreed	DATE	ddmmyyyy
Month of decision to treat	month_of_decision_to_treat_mm	Month	DATE	mm
Year of decision to treat	year_of_decision_to_treat_yyyy	Year	DATE	yyyy
Start Date of Regimen	start_date_of_regimen	This is the first administration date of the first cycle of a regimen.	DATE	ddmmyyyy
Month start for drug regimen	month_start_for_drug_regimen	month	DATE	mm
Year of start for drug regimen	year_start_for_drug_regimen	Year	DATE	yyyy
Clinical Trial	clinical_trial	Indicates whether an individual episode of care	TEXT	01 = PATIENT is taking

Indicator		was delivered as part of a clinical trial.		part in a CLINICAL TRIAL 02 = PATIENT is not taking part in a CLINICAL TRIAL 99 = Not Known

4. Cycle

Column description	*Column Name*	*Details*	*Field Type*	*Valid Content*
CR patient Identifier	e_cr_patid	Unique patient identifier based on NCRAS data. In some cases, the same person may have multiple patient IDs. Patient IDs will be retained even after two patient records are found to be the same person.	ID	Number
Pseudonymised Tumour Identifier	e_pseudo_merged_tumour_id	Tumour Identifier	NUMBER	Max length 6 characters
Pseudonymised Regimen identifier	e_pseudo_merged_regimen_id	Regimen Identifier	NUMBER	Max length 6 characters
Pseudonymised Cycle Identifier	e_pseudo_merged_cycle_id	Unique identifier for a Cycle within a Regimen (used to link to drug detail table).	NUMBER	Max length 7 characters
Cycle Identifier	cycle_number	Number of cycle	NUMBER	
Start Date (Cycle)	start_date_of_cycle	Date of the first drug administration in each Cycle.	DATE	ddmmyyyy
Month of start date of cycle	month_of_start_date_of_cycle	Month	DATE	mm
Year of start date of cycle	year_of_start_date_of_cycle	Year	DATE	yyyy
Person Weight	weight_at_start_of_cycle	Weight in kilograms at the start of the Cycle. Note that the completion of this variable is optional as required for local purposes.	NUMBER	Range: 1 - 985 Recorded to 3 d.p.
Performance Status	performance_status_at_start_of_cycle_clean	Performance status at start of cycle (adult)	TEXT	0 = Able to carry out all normal activity without restriction 1 = Restricted in physically strenuous activity, but able to walk and do light work 2 = Able to walk and capable of all self care, but unable to carry out any

				work. Up and about more than 50% of waking hours 3 = Capable of only limited self care, confined to bed or chair more than 50% of waking hours 4 = Completely disabled. Cannot carry on any self care. Totally confined to bed or chair

5. Drug detail

Column description	*Column Name*	*Details*	*Field Type*	*Valid Content*
CR patient Identifier	e_cr_patid	Unique patient identifier based on NCRAS data. In some cases, the same person may have multiple patient IDs. Patient IDs will be retained even after two patient records are found to be the same person.	ID	Number
Pseudonymised Tumour Identifier	e_pseudo_merged_tumour_id	Tumour Identifier	NUMBER	Max length 6 characters
Pseudonymised Regimen identifier	e_pseudo_merged_regimen_id	Regimen Identifier	NUMBER	Max length 6 characters
Pseudonymised Systemic Anti-Cancer Therapy Cycle Identifier	e_pseudo_merged_cycle_id	Cycle Identifier is a unique identifier for an Anti-Cancer Drug Cycle within an Anti-Cancer Drug Regimen.	NUMBER	Max length 7 characters
Drug Detail identifier	e_pseudo_merged_drug_detail_id	Drug Detail Identifier	NUMBER	Max length 7 characters
Drug Analysis Grouping	drug_group	This is a cleaned version of the drug name. Note includes 'Not Chemo', 'Not matched'.	TEXT	Max Length 30 characters
Actual Dose	actual_dose_per_administration	The actual chemotherapy dose can be given in milligrams (mg) or other applicable units, for example, 400 milligrams, 200 units, 1.5 grams etc.	NUMBER	Range: Min -1, Max 6575
The unit of measure used for each SACT drug administration	admin_measure_per_actual_dose	Units of measurements used for each administration in a SACT cycle. This descriptive field should be provided in combination with Actual_Dose_Per_Administration. Not available prior to 2018.	NUMBER	Max length 2 characters 01=mg 02=Mcg 03=g 04=Units 05=Cells 06=x10^6 PFU 07=x10^8 PFU 98=Other 99=Unknown
SACT drug route of Administration	administration_route	Route of administration for SACT drug	NUMBER	01 =Intravenous 02 = Oral

				03 = Intrathecal 04 = Intramuscular 05 = Subcutaneous 06 = Intraarterial 07 = Intraperitoneal 08 = Other intracavity Intracavernous 09 = Intravesical (Intra-Vesicular) 10 = Intratumour / Intralesional 11 = Cutaneous (Topical) 12 = Intradermal 13 = Intratumour 14 = Intralesional 98 = Other
SACT Administration Date	administration_date	Date on which the anti-cancer drug was administered to a patient, an infusion commenced, or an oral drug was initially dispensed to the patient.	DATE	ddmmyyyy

6. Outcome

Column description	*Column Name*	*Details*	*Field Type*	*Valid Content*
CR patient Identifier	e_cr_patid	Unique patient identifier based on NCRAS data. In some cases, the same person may have multiple patient IDs. Patient IDs will be retained even after two patient records are found to be the same person.	ID	Number
Pseudonymised Tumour Identifier	e_pseudo_merged_tumour_id	Tumour Identifier	NUMBER	Max length 6 characters
Pseudonymised Regimen identifier	e_pseudo_merged_regimen_id	Regimen Identifier	NUMBER	Max length 6 characters
Pseudonymised Outcome Identifier	e_pseudo_merged_outcome_id	Outcome Identifier	NUMBER	Max length 6 characters
Regimen modification indicator (dose reduction)	regimen_modification_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.	TEXT	Y = Yes, N = No
Regimen outcome summary – curative (not completed as planned) reason	Regoutsum_cur_not_com_plan	Reason for not completing regimen as planned	TEXT	1 - Progressive/recurrent cancer 2 - Toxicity 3 - Death 4 - Patient choice 5 - Other
Regimen outcome summary – non-curative	Regoutsum_non_curat	An indicator of whether the patient benefited from the non-curative treatment	TEXT	Y/N
Regimen outcome summary - toxicity	Regoutsum_toxic	An indicator of the toxicity of the regimen	TEXT	Y/N
Regimen outcome summary – curative (completed as planned)	Regoutsum_cur_com_plan	An indicator of whether the patient completed the regimen as planned	TEXT	Y/N