UK Colorectal Cancer
INTELLIGENCE HUB

CORECT-R DATA
CATALOGUE V1.0

Contents

Introduction

Colorectal cancer is a major public health problem. Each year in the UK around 41,000 people are diagnosed with the disease and 16,000 die from it. Overall, survival rates are poor. Detecting and managing the illness costs the NHS over a billion pounds a year but, despite this outlay, there remain major variations in diagnosis, treatment and outcomes. In parallel, the research community invests significant resource to help our understanding of the aetiology of the disease as well as to develop more effective methods to detect and manage it.

High-quality data could help address these challenges and improve colorectal cancer outcomes. This is because good intelligence underpins patient choice, helping individuals reduce the risk of disease and access the best care. It also helps identify and quantify inequalities, improves the cost-effectiveness and quality of services, and supports cancer research. Unfortunately, the availability of such high-quality cancer data and intelligence has been limited. Cancer Research UK has, therefore, funded the UK Colorectal Cancer Intelligence Hub to help drive improvements in the colorectal cancer data available for research. To achieve this the Hub is developing a resource known as the COloRECTal cancer data Repository or CORECT-R in which as many datasets as possible that are relevant to colorectal cancer from across the UK are robustly linked and made accessible via a secure Trusted Research Environment (TRE) where researchers can analyse these data.

This document details the current data available within CORECT-R. It details all the data attributes in all the data tables incorporated into the resource. It is intended to help support and inform users of the available data within CORECT-R. It will continually be updated as more information is incorporated into the resource.

Summary of CORECT-R content

In each of the four UK nations, and internationally, there are numerous individual datasets containing important information about colorectal cancer and its management alongside repositories of linked cancer data. There are also many research and service groups all seeking access to extracts of these datasets, undertaking bespoke linkages using different methods and, in consequence, obtaining slightly different results. There is, therefore, significant duplication of effort, and inefficiency and confusion for those wishing to make use of the intelligence. CORECT-R seeks to resolve this by providing a single point of linkage for all datasets as well as transparent processing methods leading to the export of de-identified extracts of data for analysis in a Trusted Research Environment.

At the heart of CORECT-R are national colorectal and anal datasets. These collate variables from many of the component datasets, alongside newly derived variables, to produce a population-based 'research ready' dataset for analysis. The current contents of these national datasets are available in the Data Profiles section of this document. In addition, full information on how all the derived variables in these datasets have been created are included in the 'Derived Variables' section.

There are also numerous other administrative and research datasets relevant to the disease (and used to create the National Colorectal and Anal Dataset) that can also be accessed through CORECT-R. The respective owners have previously catalogued these datasets so, rather than replicate this information, this document provides links and information which CORECT-R users can follow to find out more about the data they offer.

How to access data within CORECT-R

Full details of how to access CORECT-R data are available in the UK Colorectal Cancer Intelligence Hub's website at http://www.ndph.ox.ac.uk/CORECTR but, in summary, the process involves

1. A user contacting the UK Colorectal Cancer Intelligence Hub to discuss their project.
2. A protocol being developed, in close collaboration with the Hub team, to describe the rationale for the project, the data required, the analyses proposed and the expected outcomes.
3. The protocol being considered by the Hub's data access team and, if approved, work commencing on the project.

If you wish to use CORECT-R contact the Hub team, crchub@leeds.ac.uk

Data Profile: National Colorectal Cancer Dataset

Introduction

This dataset includes information on every colorectal cancer diagnosed in England.

The spine of the dataset is all colorectal cancers diagnosed in England identified by the National Cancer Registration and Analysis Service, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/884776/A_guide_to_NCRAS_data_and_its_availability__1_.pdf These data have then be linked to other datasets to supplement the available information for each individual's care.

This dataset has been produced in an effort to both protect the confidentiality of the colorectal cancer population whilst also increasing the ease of access to the information for researchers in an attempt support the generation of intelligence that can help improve colorectal cancer care. The dataset has the intention of combining key, quality assured information derived from across the full range of datasets available to provide robust summary patient-level data that will be sufficient for the majority of research requests made.

This National Colorectal Cancer Dataset is being continually developed and its scope extended through ongoing linkage to new datasets and the derivation of new variables that can then be incorporated into routine use.

Data Summary

Data Provider	UK Colorectal Cancer Intelligence Hub
Temporality of the data	*April 1997-December 2017*
Geographical extent	England
Data tables	1

Content

The data are provided in a single table. Some variables within the dataset are taken directly from other datasets. Others are derived using algorithms relating information from multiple datasets. The source of each variable can be determined using the key below. The algorithms used to produce any derived variables are described in the 'Derived Variables' section of this catalogue.

Variable	Cancer Registry
Variable	Hospital Episode Statistics (HES)
Variable	Cancer Patient Experience Survey (CPES)
Variable	Cancer Survivor Patient Reported Outcomes (PROMs) Survey
Variable	CORECT-R

V1.0 September 2020

Variables

Data Item	Variable name	Description of field content	Format	Further info
Identifiers				
Pseudonymised person ID	■ PERSONID	Unique identifier for each patient. Derived by the UK Colorectal Cancer Intelligence Hub.	Number	
Pseudonymised tumour ID	■ TUMOURID	Unique identifier for each tumour. Derived by the UK Colorectal Cancer Intelligence Hub.	Number	
Patient information				
Age	■ AGE	Option to derive age information in a format suitable for the research project needs. Can be single age or summarised in age bands (e.g. 5-year) and can could also include month and year of birth	Number	
Sex	■ SEX	0=Not known, 1=Male, 2=Female, 9=Not specified.	Number	
Ethnic group	■ ETHNICITY	Option to group ethnicities (e.g. white/ non-white/ unknown). Several methods of deriving ethnicity information have been used previously and can be derived as per user requirements	Text	
Socioeconomic status	■ IMD		Number	
Vital status of the patient	■ VITALSTATUS	A =Alive, D =Dead, X =Exit posting	Char(1)	
Days from diagnosis to death	■ SURV	Time in days between specified colorectal cancer diagnosis and death or censoring	Number	SURV – days from diagnosis to death
Days from another event to date to death	■ DODINT	Option to provide number of days from another event to death (e.g. days from diagnosis to death). Derived as per user requirements	Number	
Cause of death	■ COD	Option to derive summary cause of death information based on the causes of death listed on death certificates or coded by former regional registries. Derived as per user requirements	Text	
Place of death	■ POD	Option to derive place of death based on information recorded on death certificates or from Hospital Episode Statistics. Derived as per user requirements		

V1.0 September 2020

Tumour information					
Diagnosis date	■	DIAGDATE	Option to derive diagnosis date in a format suitable for the research project needs. This is likely to be in the format MM/YYYY unless permissions for access to identifiable data have been granted		
Days from another event to date to diagnosis	■		Option to provide number of days from another event to diagnosis (e.g. days from birth to diagnosis). Derived as per user requirements	Number	
Basis of diagnosis of the tumour	■	BASISOFDIAGNOSIS	0 = Death certificate, 1 = Clinical: Diagnosis made before death without (2-7), 2 = Clinical investigation, 4 = Specific tumour markers, 5 = Cytology, 6 = Histology of a metastases, 7 = Histology of a primary tumour, 9 = Unknown, e.g. PAS or HISS record only	Number	
Route to diagnosis	■	ROUTE	1=GP Referral 2=Two Week 3= Emergency Presentation 4=Other Outpatient 5=Screen Detected 6=Inpatient Elective 7=DCO 9=Unknown	Number	
Site of neoplasm (4-character code)	■	SITE_ICD10_O2	Valid 4 digit ICD-10 codes in the range C180 to C20	Char(4)	
Site of neoplasm (3-character code)	■	SITE_ICD10_O2_3CHAR	Valid 3 digit ICD-10 codes in the range C18 to C20	Char(4)	
Site of neoplasm	■	SITE_CODED	1=Right colon 2=Left colon 3=Colon, unspecified 4=Rectosigmoid 5= Rectum	Number	SITE_CODED – location of the tumour within the colorectum
Morphology	■	MORPH_CODED	The original five digit ICD03 morphology information captured by NCRAS	Char(5)	
Morphology Group	■	USE_MORPH	A grouped morphology variable derived by grouping the detailed morphology information available in NCRAS to provide the more useable histology types of adenocarcinoma, etc. For many analyses it may be appropriate to limit cases to exclude rare morphological sub types with non-standard treatment or unusual behaviour.	Number	USE_MORPH – morphology group
Summary stage	■	USESTAGE	A summary stage, that combines information from the multiple staging variables available in NCRAS to maximise the number of cases with a stage classification	Number	USESTAGE – summary stage
Staging information	■		Option to derive staging information from any of the data items available in the CORECT-R tables (likely to be predominantly NCRAS AV Tumour). This can include individual components of TNM or any other information that can be reported alone or combined as demanded by the research project. Derived as per user requirements		

Comorbidity information				
Comorbidity information	■	Option to derive comorbidity information using diagnosis information stored in the HES data. Derived as per user requirements. This could include using a previously published application of the Charlson, Elixhauser or C3 score or aDCSI information, or bespoke algorithms as demanded by the research project.	Number	COMORBIDITY INFORMATION
Primary procedure information				
Primary procedure	■ PROCTYPE	Primary procedure for this individual 1=Major resection 2=Minor resection 3=Bypass 4=Stoma 5=Stent 6=No surgery 7= No link to HES so no information available	Number	PROCTYPE – primary procedure
Primary procedure OPCS4 code	■ OPCSC	The OPCS4 code indicating the exact procedure determined as the primary event	Char(4)	
Date of primary procedure	■ OPYEAR	MMYYYY	Number	
Method of admission to hospital for primary procedure	■ ADMITMETH	Derived from the HES variable ADMISSION METHOD. 0=Elective 1=Emergency 99=Unknown	Number	ADMITMETH – method of admission to hospital for primary procedure
Location patient admitted from	■ ADMITLOC	1=Unsupervised accommodation, no care provided 2=Hospital 3=Supervised accommodation, health care is provided 4=Supervised accommodation, no care provided 99=Unknown	Number	
Urgency of surgery	■ SURGURG	0=Elective 1=Emergency	Number	SURGURG – urgency of surgery
Destination patient discharged to after primary procedure	■ DISCHLOC	1=Unsupervised accommodation, no care provided 2=Hospital 3=Supervised accommodation, health care is provided 4=Supervised accommodation, no care provided 9=Patient died 99=Unknown	Number	
Days between diagnosis and primary procedure	■ DIAGINT	Option to provide number of days from another event to diagnosis (e.g. days from birth to diagnosis). Derived as per user requirements	Number	DIAGINT – days between diagnosis and primary procedure
Interval between primary procedure and	■ DEATHINT	Option to provide number of days from another event to date of death (e.g. days from birth to diagnosis). Derived as per user requirements	Number	DEATHINT – days between primary

death or when censored				procedure and death
Length of stay in days in the spell of care associated with the primary procedure	■ LOS	Total length of stay in hospital associated with the primary procedure	Number	LOS – length of stay
Length of post-operative stay in hospital	■ POLOS	Length of post primary procedure stay in hospital in days	Number	POLOS – post-operative length of stay
Emergency readmission within 30 days of discharge following primary procedure	■ READMI	0=No 1=Yes	Number	READMI – emergency readmission within 30 days of discharge
Death within 30 days of primary procedure	■ DM30	0=Alive 1=Dead	Number	DM30 – death within 30-days of primary procedure
Death within 90 days of primary procedure	■ DM90	0=Alive 1=Dead	Number	DM90 – death within 90-days of primary procedure
Death during hospital stay of primary procedure	■ HOSPDEATH	0=No 1=Yes	Number	HOSPDEATH – death during hospital stay for primary procedure
Trust of primary procedure	■ HOSPBCI		Char(5)	HOSPBCI – Trust where primary procedure took place
MDT of primary procedure	■ MDT		Char(5)	MDT – Multi-disciplinary team responsible for

					primary procedure
Further information if primary procedure is a major resection					
Procedure group	■	PROCNAME	1=Total excision of colon & rectum 2=Total excision of colon 3=Subtotal excision of colon 4=Extended right hemicolectomy 5=Right hemicolectomy 6=Transverse colectomy 7=Left hemicolectomy 8=Sigmoid colectomy 9=Other colectomy 10=APE 11=Anterior resection 12=Hartmann's procedure 13=Pelvic exenteration 14=Other rectal	Number	
Stoma opened at major resection	■	STOMA	0=No 1=Yes	Number	STOMA – stoma opened during major surgical resection
Stoma still open at 18 months following major resection	■	STOMA18	0=Closed 1=Open	Number	STOMA18 – stoma open at 18 months following creation
Approach to major resection	■	LAP	0=Open 1=Laparoscopic	Number	APPROACH – approach to major surgical resection
Converted laparoscopic procedures	■	CONVERT	0=Not converted 1=Converted 9=Not laparoscopic	Number	CONVERT – procedures converted from laparoscopic to open
Type of laparoscopic procedure	■	APPTYPE	0=Standard laparoscopic 1=Robotic laparoscopic	Number	APPTYPE – type of laparoscopic procedure
Consultant listed as overseeing the episode of care associated with the major resection	■	CONS		Char(9)	
ACPGBI membership status of Consultant undertaking major resection	■	ACPGBI	0 – Non-member of the ACPGBI 1 – Member of the ACPGBI	Char(1)	

9

Management geographies				
Geography of diagnosis	■ GEOGD	Option to provide any geography listed at LOCATION based on the postcode of residence of the patient at the time of diagnosis of the colorectal tumour. Derived as per user requirements		
Geography of diagnosis and treatment	■ GEOGT	Option to provide any geography listed at LOCATION based on the hospitals of diagnosis and treatment of the patient at any time in their pathway of care. Derived as per user requirements		
Travel time	■ TRAV	Option to provide road travel time for each patient between their home and relevant hospitals they attended at any point across the care pathway. Derived as per user requirements		
Travel distance	■ TRAVDIST	Option to provide road travel distance for each patient between their home and relevant hospitals they attended at any point across the care pathway Derived as per user requirements		
Oncology				
Neoadjuvant rectal cancer treatment	■ RECTALRT	0 – No radiotherapy 1 – Short course radiotherapy with immediate surgery 2 – Short course radiotherapy with delayed surgery 3 – Long course chemoradiotherapy 4 – Post-operative radiotherapy 5 – Other radiotherapy	Number	RECTALRT – neoadjuvant rectal cancer treatment
Adjuvant chemotherapy	■ ADJCT	0=No 1=Yes	Number	ADJCT – adjuvant chemotherapy
Type of chemotherapy	■ ADJCT_TYPE	1=single agent 2=combination agent	Number	ADJCT_TYPE – type of adjuvant chemotherapy
Patient reported outcomes & experience				
Patient reported outcomes	■	Option to provide information from the 2013 colorectal cancer PROMs survey. Derived as per user requirements. These results cannot be linked to the registry or HES data but are available as an independent series of data items		PATIENT REPORTED OUTCOMES (PROMs)
Cancer patient experience survey	■	Option to provide information from the Cancer Patient Experience survey. Derived as per user requirements. These results cannot be linked to the registry or HES data but are available as an independent series of data items		

Data Profile: National Anal Cancer Dataset

Introduction

This dataset includes information on every anal cancer diagnosed in England.

The spine of the dataset is all anal cancers diagnosed in England identified by the National Cancer Registration and Analysis Service (Dataset link). These data have then been linked to other datasets to supplement the available information for each individuals care.

This dataset has been produced in an effort to both protect the confidentiality of the anal cancer population whilst also increasing the ease of access to the information for researchers in an attempt support the generation of intelligence that can help improve anal cancer care. The dataset has the intention of combining key, quality assured, information derived from across the full range of datasets available to provide robust summary patient-level data that will be sufficient for the majority of requests made.

This National Anal Cancer Dataset is being continually developed and its scope extended through ongoing linkage to new datasets and the derivation of new variables that can then be incorporated into routine use.

Data Summary

Data Provider	UK Colorectal Cancer Intelligence Hub
Temporality of the data	*April 1997 – December 2017*
Geographical extent	England
Data tables	1

Content

The data are provided in a single table. Some variables within the dataset are taken directly from other datasets. Others are derived using algorithms relating information from multiple datasets. The source of each variable can be determined using the key below. The algorithms used to produce any derived variables are described in the Derivations section of the catalogue.

Variables highlighted in red contain highly sensitive data. These data items can be requested but additional permissions will be required to enable their release for research.

Variable	NCRAS
Variable	HES
Variable	CORECT-R
Variable	RTDS
Variable	SACT

Variables

Data Item	Variable name	Description of field content	Format	
Identifiers				
Pseudonymised person ID	■ PERSONID	Unique identifier for each patient. Derived by the UK Colorectal Cancer Intelligence Hub.	Number	
Pseudonymised tumour ID	■ TUMOURID	Unique identifier for each tumour. Derived by the UK Colorectal Cancer Intelligence Hub.	Number	
Person information				
Age	■ AGE	Option to derive age information in a format suitable for the research project needs. Can be single age or summarised in age bands (e.g. 5-year) and can could also include month and year of birth	Number	
Sex	■ SEX	0=Not known, 1=Male, 2=Female, 9=Not specified.	Number	
Ethnic group	■ ETHNICITY	Option to group ethnicities (e.g. white/ non-white/ unknown). Several methods of deriving ethnicity information have been used previously and can be derived as per user requirements	Text	
Socioeconomic status	■ IMD		Number	
Date of death	■ DOD	Option to derive date of death information in a format suitable for the research project needs. Derived as per user requirement		
Vital status of the patient	■ VITALSTATUS	A =Alive, D =Dead, X =Exit posting	Char(1)	
Days from diagnosis to death	■ SURV	Time in days between specified colorectal cancer diagnosis and death or censoring	Number	SURV – days from diagnosis to death
Days from another event to date to death	■ DODINT	Option to provide number of days from another event to death (e.g. days from diagnosis to death). Derived as per user requirements	Number	
Cause of death	■ COD	Option to derive summary cause of death information based on the causes of death listed on death certificates or coded by former regional registries. Derived as per user requirements		
Place of death	■ POD	Option to derive place of death based on information recorded on death certificates or from Hospital Episode Statistics. Derived as per user requirements		
Tumour information				
Diagnosis date	■ DIAGDATE	Option to derive diagnosis date in a format suitable for the research project needs. Derived as per user requirements		

V1.0 September 2020

Days from another event to date to diagnosis	■ DIAGINT	Option to provide number of days from another event to diagnosis (e.g. days from birth to diagnosis). Derived as per user requirements		
Basis of diagnosis of the tumour	■ BASISOFDIAGNOSIS	0 = Death certificate, 1 = Clinical: Diagnosis made before death without (2-7), 2 = Clinical investigation, 4 = Specific tumour markers, 5 = Cytology, 6 = Histology of a metastases, 7 = Histology of a primary tumour, 9 = Unknown, e.g. PAS or HISS record only		
Route to diagnosis	■ ROUTE	1=GP Referral 2=Two Week 3= Emergency Presentation 4=Other Outpatient 5=Screen Detected 6=Inpatient Elective 7=DCO 9=Unknown		
Site of neoplasm (4-character code)	■ SITE_ICD10_O2	Valid 4 digit ICD-10 codes for site of tumour	Char(4)	
Morphology	■ MORPH_CODED	The original five digit ICD03 morphology information captured by NCRAS	Char(5)	
Morphology Group	■ USEMORPH_ANAL	A grouped morphology variable derived by grouping the detailed morphology information available in NCRAS to provide the more useable histology types. For many analyses it may be appropriate to limit cases to exclude rare morphological sub types with non-standard treatment or unusual behaviour.		USEMORPH_ANAL
Summary stage	■ USESTAGE_ANAL	A summary stage, that combines information from the multiple staging variables available in NCRAS to maximise the number of cases with a stage classification		USESTAGE_ANAL
Staging information	■	Option to derive staging information from any of the data items available in the CORECT-R tables (likely to be predominantly NCRAS AV Tumour). This can include individual components of TNM or any other information that can be reported alone or combined as demanded by the research project. Derived as per user requirements		
Comorbidity information				
Comorbidity information	■	Option to derive comorbidity information using diagnosis information stored in the HES data. Derived as per user requirements. This could include using a previously published application of the Charlson, Elixhauser or C3 score or aDCSI information, or bespoke algorithms as demanded by the research project.	Number	COMORBIDITY INFORMATION
Surgical information				
Abdominoperineal resection	■ APE	1= APE 0= No APE	Number	APE
Interval	OP_INT	Interval (in days) between anal cancer diagnosis and APE	Number	

Surgical information	■	Option to provide surgical treatment information. Derived as per user requirements		
Oncology				
Radiotherapy information	■	Option to provide radiotherapy treatment information. Derived as per user requirements		
Chemotherapy information	■	Option to provide chemotherapy treatment information. Derived as per user requirements		
Links to other datasets				

Data Profiles:

National Cancer Registration & Analysis Service Colorectal Cancer Data

This dataset includes information on every colorectal and anal cancer diagnosed in England. These data are captured by the National Cancer Registration and Analysis Service (NCRAS) within Public Health England. NCRAS achieves comprehensive registration of all registerable tumours by bringing together data from more than 500 local and regional datasets to build a picture of all individuals care pathways from diagnosis onwards. Although the NCRAS information in CORECT-R is limited to only colorectal and anal cancers, the organisation captures data on all cancers and this is accessible via application through the Public Health England Office for Data Release.

Data Summary

Data Provider	**NCRAS**
Temporality of the data	*1995 to 2017*
Geographical extent	England
Data tables	Patient
	Tumour
	Treatment
Further information and data dictionary	https://www.gov.uk/government/publications/accessing-public-health-england-data

Hospital Episode Statistics

The Hospital Episode Statistics (HES) dataset includes information on every hospital attendance in England. These data are collected during a patient's time at hospital to allow these organisations to be paid for the care they deliver. HES data are designed to enable secondary use, i.e. non-clinical purposes.

The data are provided in three tables. The largest, the inpatient table, relates to episodes of care that occurred following a hospital or day case admission. The outpatient table captures outpatient attendances and accident and emergency all attendances within an accident and emergency unit within the English NHS. The data within these tables are linked, at a person level, to other datasets within CORECT-R

The HES information in CORECT-R is limited to only those episodes of care relevant to individuals who have had a diagnosis of colorectal and anal cancers recorded by the NCRAS and/or who have a diagnosis of colorectal or anal cancer recorded in HES. All episodes of care for such individuals are available. The HES dataset does, however, contain information on all hospital attendances in England and this information is available on application to NHS Digitals Data Access Request Service (https://digital.nhs.uk/services/data-access-request-service-dars).

Data Summary

Data Provider	**NHS Digital**
Temporality of the data	*April 1997 to December 2019*
Geographical extent	England
Data tables	Inpatient
	Outpatient
	Accident & Emergency
Further information	https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics
Data dictionaries	https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary

Systemic Anti-Cancer Therapy Dataset (SACT)

The Systemic Anti-Cancer Therapy (SACT) data set collects clinical management on patients receiving cancer chemotherapy in or funded by the NHS in England. The dataset has been designed to collect information on all drug treatments with an anti-cancer effect, in all treatment settings, including traditional cytotoxic chemotherapy and all newer agents.

The SACT information in CORECT-R is limited to only those records available for individuals who have had a diagnosis of colorectal and anal cancers recorded by the NCRAS. Linked data are only available for cases diagnosed from 2014 onwards.

Data Summary

Data Provider	**NCRAS**
Temporality of the data	*April 2012 - December 2017*
Geographical extent	England
Data tables	Demographics and consultant
	Clinical status
	Regimen
	Cycle
	Drug details
	Outcome
Further information	www.chemodataset.nhs.uk
Data dictionary	https://www.datadictionary.nhs.uk/data_dictionary/messages/clinical_data_sets/data_sets/systemic_anti-cancer_therapy_data_set_fr.asp

National Radiotherapy Dataset (RTDS)

The Radiotherapy Data Set (RTDS) is submitted by all NHS Acute Trust providers of radiotherapy services in England and is used to provide intelligence for service planning, commissioning, clinical practice and research and the operational provision of radiotherapy services across England.

The National Clinical Analysis and Specialised Applications Team (NATCANSAT) based at The Clatterbridge Cancer Centre NHS Foundation Trust were responsible for the management and delivery of RTDS from April 2009 until the 1st April 2016 when Public Health England took over full responsibility for capturing and curating RTDS.

The RTDS includes information on teletherapy, brachytherapy given using automated remote afterloading machines and all other brachytherapy given for the treatment of malignant disease delivered in England to patients in NHS facilities, or in private facilities where delivery was funded by the NHS.

The RTDS does not include brachytherapy delivered using methods other than automated remote afterloading, for the treatment of non-malignant disease, radiotherapy delivered using unsealed sources, radiotherapy delivered outside of England, radiotherapy delivered in a non-NHS setting and not funded by the NHS, non-therapeutic exposures delivered using a radiotherapy machines (e.g. imaging) and non-patient exposures (e.g. dosimetry exposures, blood or tissues, animals).

The RTDS information in CORECT-R is limited to only those records available for individuals who have had a diagnosis of colorectal and anal cancers recorded by the NCRAS.

Data Summary

Data Provider	**The National Clinical Analysis and Specialised Applications Team (NATCANSAT)** between 1st April 2009 and 31st March 2016. **National Cancer Registration (NCRS)** at Public Health England (PHE) from 1st April 2016
Temporality of the data	*April 2009 – December 2017*
Geographical extent	England
Further information	http://www.ncin.org.uk/collecting_and_using_data/rtds
Data dictionary	https://www.datadictionary.nhs.uk/data_dictionary/messages/clinical_data_sets/data_sets/radiotherapy_data_set_fr.asp?shownav=1

Cancer Patient Experience Survey (CPES)

In its first 6 years, the National Cancer Patient Experience Survey (CPES) was commissioned by NHS England through Quality Health and was a survey sent out to all adult cancer patients (aged 16 and over) with a primary diagnosis of cancer who have been admitted to an acute or specialist NHS Trust in England providing adult cancer services as inpatients or day cases, and discharged within a specified three month sampling period each year. The survey aimed to collect information from patients about their cancer journey from their initial GP visit prior to diagnosis, through diagnosis and treatment and to the ongoing management of their cancer.

The CPES questionnaire used has varied slightly across these different time periods with the loss and addition of questions in different versions.

The CPES information in CORECT-R is limited to only those survey participants who had a diagnosis of colorectal or anal cancers recorded by the NCRAS or HES.

Data Summary

Data Provider	**Quality Health** (2010-2015) **Picker** (2016-2018)
Temporality of the data	2010: Patients discharged between 01/01/2010 and 31/03/2010
	2011: Patients discharged between 01/09/2011 and 30/11/2011
	2012: Patients discharged between 01/09/2012 and 30/09/2012
	2013: Patients discharged between 01/09/2013 and 30/11/2013
	2013: Patients discharged between 01/09/2014 and 30/11/2014
	2015: Patients discharged between 01/04/2015 and 30/06/2015
	2016: Patients discharged between 01/04/2016 and 30/06/2016
	2017:Patients discharged between 01/04/2017 and 30/06/2017
	2018:Patients discharged between 01/04/2018 and 30/06/2018
Geographical extent	England
Further information	https://www.ncpes.co.uk/
Data dictionary	https://www.ncpes.co.uk/reports

Cancer Survivor Patient Reported Outcomes (PROMs) Survey

The Quality of Life of Colorectal Cancer Survivors in England national survey (2013) was commissioned by NHS England as part of the National Cancer Survivorship Initiative (NCSI) and was conducted by Picker Institute Europe in conjunction with NCRAS. The survey measured overall and cancer-specific quality of life in patients still alive 12-36 months after a diagnosis of colorectal cancer. The data available in CORECT-R relate to the responses from the colorectal cancer patients that completed the survey.

Outcome questions in the survey are made up of three instruments: the EQ-5D-5L (Euroqol 5-level), FACT (Functional Assessment of Cancer Therapy) items and SDI (Social Difficulties Inventory).

Data Summary

Data Provider	**Department of Health**
Temporality of the data	*Patients diagnosed in 2010 and 2011, surveyed in 2013*
Geographical extent	England
Further information	https://www.england.nhs.uk/wp-content/uploads/2015/03/colorectal-cancer-proms-report-140314.pdf

Derived Variables:

The following section provides information about derived variables contained within the National Colorectal Cancer Dataset and the National Anal Cancer Dataset. Variables are grouped according to themes;

1. Patient information
2. Tumour information
3. Comorbidity information
4. Primary procedure
5. Major resection
6. Oncology
7. Patient Reported Outcomes (PROMs) (available in isolation only)
8. Anal cancer (containing information about variables which are present in the National Anal Cancer Dataset only)

National Colorectal Cancer Dataset

PATIENT INFORMATION

Days from diagnosis to death - SURV

Data item:	Days from diagnosis to death
Field name:	SURV
Format:	Number
Description of field content:	Number of days
Valid event date range:	April 1997 – December 2017
Data source:	NCRAS
Generate by:	UK Colorectal Cancer Intelligence Hub
Associated variables:	VITALSTATUS

Rules for derivation

The variable SURV contains the time period from diagnosis of the relevant colorectal tumour to death for all patients with a date of death recorded. For patients with no date of death a census date is used. A random offset of between 1 and 5 days is applied to the census date to create a randomised census date for each patient.

Comments

In order to create pseudonymised data a random offset of between 1 and 5 days is applied to the census date for the creation of this variable.

TUMOUR INFORMATION

Location of the tumour within the colorectum – SITE_CODED

Data item:	Location of the tumour
Field name:	SITE_CODED
Format:	Number
Description of field content:	1 – Right colon 2 – Left colon 3 – Colon, unspecified 4 – Rectosigmoid 5 – Rectum
Valid event date range:	April 1997 – December 2017
Data source:	NCRAS
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

The variable SITE_CODED is used to specify the location of the tumour within the colon, rectosigmoid or rectum. The classification is based on the ICD10 code for the tumour recorded within the cancer registry data.

Morphology group – USE_MORPH

Data item:	Morphology group of the tumour
Field name:	USE_MORPH
Format:	Number
Description of field content:	1 – Adenocarcinoma 2 – Squamous 3 – Melanoma 4 – Cancer/carcinoma NOS 5 – Mixed 6 – Neuroendocrine 7 – Sarcoma 9 - Other
Valid event date range:	April 1997 – December 2017
Data source:	NCRAS
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

The variable USE_MORPH is used to classify the morphology of the tumour within the colon, rectosigmoid or rectum. The classification is based on the ICDO2 code for the tumour recorded within the cancer registry data. Morphologies which don't fall into categories 1-7 are classified as other.

Data quality

This variable includes only tumours of the colon, rectosigmoid and rectum (C18-C20). Morphology groupings for anal tumours can be found in USEMORPH_ANAL.

Summary stage - USESTAGE

Data item:	Summary stage
Field name:	USESTAGE
Format:	Number
Description of field content:	1 – I 2 – II 3 – III 4 – IV 9 - Unknown
Valid event date range:	April 1997 – December 2017
Data source:	NCRAS
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

The USESTAGE variable classifies the stage of disease using multiple sources within the NCRAS data to provide a single source of information.

Tumours are classified using the STAGEB field from NCRAS in the first instance. Tumours which are not allocated to a stage following this are then classified using the DUKES stage field from NCRAS. For tumours where no classification is possible following these steps the number of positive nodes are examined (NODESP from NCRAS), where this is positive tumours are assigned to the stage III group. Any remaining tumours are classified using the METDS field where this is complete. In all cases where this is positive tumours are allocated to the stage IV group, regardless of previous staging information. Any remaining tumours are assigned an unknown stage.

NCRAS staging variable				CORECT-R code	CORECT-R description
STAGEB	DUKES	NODESP	METSD		
Step 1	Step 2	Step 3	Step 4		
1				1	**I**
1A					
1A2					
1B					
1C					
1E					
1S					
2				2	**II**
2A1					
2A					
2C					
2E					
2S					
3				3	**III**
3A					
3B					
3C					
3E					
3S					
4				4	**IV**
4A					
4B					
4C					
4S					
Not completed	A			1	**I**
	B			2	**II**
	C			3	**III**
	C1				
	C2				
	D			4	**IV**
Not completed		>0		3	**III**
Not completed			1	4	**IV**
Not completed				99	**Unknown**

COMORBIDITY INFORMATION

Information below pertains to the predefined comorbidity measures which are included in the National Colorectal Cancer Dataset and the National Anal Cancer Dataset. Additional measures can be derived as per user requirements and the codes included and timeframe covered can be adapted accordingly.

Charlson comorbidity score - CHARLSON

Data item:	Charlson score
Field name:	Charlson
Format:	Number
Description of field content:	Contains the Charlson score derived for each individual and tumour.
Valid event date range:	April 1997 to December 2017
Data source:	Linked NCRAS data and HES
Generated by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable is derived from ICD10 codes for diagnoses reported in HES data records (inpatient and outpatient) for spells within the year preceding the diagnosis of the colorectal tumour in question and is aggregated to produce a Charlson score for individuals.

Previous cancer diagnoses and metastatic disease information are obtained from the NRCAS data.

Where there are multiple occurrences within the same group, only the most recently recorded comorbidity is included. Only the total score is provided.

Elixhauser comorbidity score - ELIX

Data item:	Elixhauser score
Field name:	ELIX
Format:	Number
Description of field content:	Contains the Elixhauser score derived for each individual and tumour.
Valid event date range:	April 1997 to December 2017
Data source:	NCRAS data
Generated by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable is derived from ICD10 codes for diagnoses reported in HES data records (inpatient and outpatient) for consultant episodes starting within 4 weeks after the tumour diagnosis date or within five years before tumour diagnosis date, i.e. over a 61 month period. Where there are

multiple occurrences within the same group, only the most recently recorded comorbidity is included. Each condition has a coefficient estimate assigned to it. Previous cancer diagnoses and metastatic disease information are obtained from the NRCAS data. These coefficient estimates are totalled to get the overall score for each patient.

C3 comorbidity score – C3

Data item:	C3 score
Field name:	C3
Format:	Number
Description of field content:	Contains the C3 score derived for each individual and tumour.
Valid event date range:	April 1997 to December 2017
Data source:	Linked NCRAS data and HES
Generated by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable is derived from ICD10 codes for diagnoses reported in HES data records (inpatient and outpatient) for consultant episodes starting within 4 weeks after the tumour diagnosis date or within five years before tumour diagnosis date, i.e. over a 61 month period. Where there are multiple occurrences within the same group, only the most recently recorded comorbidity is included. Each condition has a coefficient estimate assigned to it. Previous cancer diagnoses and metastatic disease information are obtained from the NRCAS data. These coefficient estimates are totalled to get the overall score for each patient.

Data quality

Anaemia, epilepsy, chronic viral hepatitis, intestinal disorders and venous insufficiency (codes in the table below) are excluded from the C3 measure due to a prevalence of less than 0.5% or being closely related to colorectal cancer.

Adapted diabetes complications index score - aDCSI

Data item:	aDCSI score
Field name:	ADCSI
Format:	Number
Description of field content:	Contains the adapted diabetes complications severity index score derived for each individual with a recorded diabetes ICD10 code (E10-E14).
Valid event date range:	April 1997 to December 2017
Data source:	Linked NCRAS data and HES
Generated by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable is derived from ICD10 codes for diagnoses reported in HES data records for consultant episodes in the six years preceding the relevant colorectal cancer diagnosis. Where there are multiple occurrences within the same group, only the most recently recorded complication is included. Each condition has a score assigned to it. If an individual has both a condition scored as moderate (score of 1) and severe (score of 2) within a single category, e.g. ocular, only the severe is included. These scores are totalled to get the overall score for each patient. The maximum possible score for any one individual is 12.

Data quality

The score is only calculated for those with a diagnosis of diabetes (ICD10 code E10-E14) recorded in the HES data in the six years prior to the diagnosis of colorectal cancer, meaning that for the majority of individuals this field will be blank.

PRIMARY PROCEDURE INFORMATION

Primary procedure - PROCTYPE

Data item:	Primary procedure
Field name:	PROCTYPE
Format:	Number
Description of field content:	1=Major resection 2=Minor resection 3=Bypass 4=Stoma 5=Stent 6=No surgery 7= No link to HES so no information available.
Valid event date range:	April 1997 – December 2017
Data source:	Linked NCRAS and HES
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

The variable PROCTYPE is used to indicate the principle surgical treatment for an individual. This variable is defined using the following rules.

1. If an individual undergoes a major resection within the month prior or up to a year after diagnosis then this is their primary procedure. If more than one major resection is recorded then the one closest to diagnosis is selected. If there is more than one major resection recorded on this day then the most extensive operation for that tumour site is retained

2. If no major resection has been identified for an individual then it is determined if, instead, a minor resection took place within the month prior or up to a year after diagnosis and, if so, then this becomes their primary procedure. If more than one minor resection is recorded then the one closest to diagnosis is selected. If there is more than one minor resection recorded on this day then the most extensive procedure is retained.

3. If no major or minor resection has been identified for an individual then the first occurring bypass or stoma opening procedure, again within a month prior to or a year after diagnosis is selected.

4. If no major or minor resection and no bypass or stoma have been undertaken then insertion of a stent is looked for again in the month prior to or up to a year after the date of diagnosis.

5. If none of the above procedures can be found for an individual then the person is designated as having NO NHS SURGERY.

6. If an individual has not attended a hospital Trust with a colorectal MDT within this period but they have attended a hospital without an MDT then their MDT is entered as NO CANCER TEAM.

Method of admission to hospital for primary procedure - ADMITMETH

Data item:	Admission method
Field name:	ADMITMETH
Format:	Number
Description of field content:	0 – Elective
	1 – Emergency
	99 - Unknown
Valid event date range:	April 1997 to December 2017
Data source:	HES data
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable is derived from the HES data variable[1] that details the nature of the patient's admission for the spell containing the surgical procedure or admission in question.

Urgency of surgery - SURGURG

Data item:	Urgency of surgery
Field name:	SURGURG
Format:	Number
Description of field content:	0 – Elective
	1 – Emergency
Valid event date range:	April 1997 – December 2017
Data source:	HES
Generate by:	UK Colorectal Cancer Intelligence Hub
Associated variables:	ADMITMETH

Rules for derivation

If the surgical intervention was undertaken within two days of an emergency admission the procedure was classified as an emergency procedure.

All other procedure episodes were classified as elective.

---

[1] https://www.datadictionary.nhs.uk/data_dictionary/attributes/a/add/admission_method_de.asp

Days between diagnosis and primary procedure - DIAGINT

Data item:	Days between diagnosis and surgical procedure
Field name:	DIAGINT
Format:	Number
Description of field content:	Number of days between date of diagnosis and surgical procedure
Valid event date range:	April 1997 to December 2017
Data source:	Linked NCRAS and HES data
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable indicates the number of days between the diagnosis of the colorectal cancer in question and the primary surgical procedure. The date of surgical procedure is identified from HES and the date of diagnosis from the NCRAS data.

Days between primary procedure and death - DEATHINT

Data item:	Days between surgical procedure and death
Field name:	DEATHINT
Format:	Number
Description of field content:	Number of days between date of surgical procedure and death
Valid event date range:	April 1997 to December 2017
Data source:	Linked NCRAS and HES data
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable indicates the number of days between the surgical procedure in question and death. The date of surgical procedure is identified from HES and the date of death from the NCRAS data. Where no procedure date was recorded it was assumed to be the date of admission.

Length of stay - LOS

Data item:	Length of stay
Field name:	LOS
Format:	Number
Description of field content:	Number of days as an inpatient between admission and discharge during the spell in which the primary procedure took place.
Valid event date range:	April 1997 to December 2017
Data source:	Linked NCRAS and HES data
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable indicates the number of days between the admission and either the date of death or discharge date for a hospital spell during which a surgical procedure occurred. Admission date and discharge date were obtained from HES data and date of death was obtained from NCRAS.

Post-operative length of stay - POLOS

Data item:	Length of stay
Field name:	POLOS
Format:	Number
Description of field content:	Number of days as an inpatient after a surgical procedure
Valid event date range:	April 1997 to December 2017
Data source:	Linked NCRAS and HES data
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable indicates the number of days between the surgical procedure date and either the date of death or discharge date for a hospital spell during which a surgical procedure occurred. Procedure date and discharge date were obtained from HES data and date of death was obtained from NCRAS.

Emergency readmission within 30 days of discharge - READMI

Data item:	Readmission
Field name:	READMI
Format:	Number
Description of field content:	0 – No readmission within 30 days 1 – Readmitted within 30 days
Valid event date range:	April 1997 – December 2017
Data source:	HES
Generate by:	UK Colorectal Cancer Intelligence Hub
Associated variables:	ADMITMETH – method of admission to hospital for primary procedure

Rules for derivation

This variable indicates whether an individual was readmitted as an emergency within 30 days of discharge following their primary procedure. Emergency admissions are identified using the admission method[2] variable in the HES data and categorised following the rules used for the ADMITMETH variable.

Death within 30-days of primary procedure – DM30

Data item:	30-day post-operative mortality
Field name:	DM30
Format:	Number
Description of field content:	0 – Alive 1 - Dead
Valid event date range:	April 1997 to December 2017
Data source:	Linked NCRAS and HES data
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable indicates whether a patient died within 30 days of the surgical procedure in question. This is derived from the procedure date as recorded in HES and the date of death as recorded by NCRAS.

---

[2] https://www.datadictionary.nhs.uk/data_dictionary/attributes/a/add/admission_method_de.asp

Death within 90-days of primary procedure – DM90

Data item:	90-day post-operative mortality
Field name:	DM90
Format:	Number
Description of field content:	0 – Alive 1 - Dead
Valid event date range:	April 1997 to December 2017
Data source:	Linked NCRAS and HES data
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable indicates whether a patient died within 90 days of the surgical procedure in question. This is derived from the procedure date as recorded in HES and the date of death as recorded by NCRAS.

Death during hospital stay for primary procedure - HOSPDEATH

Data item:	Death during hospital stay of primary procedure or major surgical resection
Field name:	HOSPDEATH
Format:	Number
Description of field content:	0 – Alive at end of hospital stay of primary procedure or major surgical resection 1 – Died during hospital stay of primary procedure or major surgical resection
Valid event date range:	April 1997 – December 2017
Data source:	Linked NCRAS and HES
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

The variable HOSPDEATH identifies individuals who died during the hospital admission (spell) during which their primary procedure or major surgical resection took place. The variable is defined using date of death, as recorded by NCRAS, and the discharge date, as reported in HES.

Trust responsible for specified event or procedure - HOSPBCI

Data item:	Hospital Trust responsible for the procedure or event in question
Field name:	HOSPBCI
Format:	Number
Description of field content:	Provider code for the Trust responsible for the procedure or event in question
Valid event date range:	April 1997 – December 2017
Data source:	Linked NCRAS and HES
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

The variable HOSPBCI identifies the hospital Trust responsible for the procedure in question, examples for colorectal cancer surgery and endoscopy are detailed below. Over time NHS geographies have changed with many Trusts merging or new ones being established. The variables SITETRET and PROCODE are used in combination to map all historical codes to current configurations. The data mapping tables (old to new configurations) are available on request from the UK Colorectal Cancer Intelligence Hub team.

Trusts can be classified as per user and project requirements.

Multi-disciplinary team responsible for primary procedure - MDT

Data item:	Colorectal multidisciplinary team responsible for the primary procedure
Field name:	MDT
Format:	Number
Description of field content:	Provider code for the colorectal cancer MDT responsible
Valid event date range:	April 1997 – December 2017
Data source:	Linked NCRAS and HES
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

The variable MDT identifies the multidisciplinary team responsible for the primary procedure for colorectal cancer. Over time NHS geographies have changed with many Trusts merging or new ones being established, as part of this process some Trusts have retained multiple MDTs.

The variables SITETRET and PROCODE are used in combination to map all historical codes to current configurations. The data mapping tables (old to new configurations) are available on request from the UK Colorectal Cancer Intelligence Hub team.

ADDITIONAL INFORMATION WHERE PRIMARY PROCEDURE IS A MAJOR RESECTION

Procedure group – PROCNAME

Data item:	Stoma opened at major resection
Field name:	PROCNAME
Format:	Number
Description of field content:	1 - Total excision of colon & rectum 2 - Total excision of colon 3 - Subtotal excision of colon 4 - Extended right hemicolectomy 5 - Right hemicolectomy 6 - Transverse colectomy 7 - Left hemicolectomy 8 - Sigmoid colectomy 9 - Other colectomy 10 - APE 11 - Anterior resection 12 - Hartmann's procedure 13 - Pelvic exenteration 14 - Other rectal
Valid event date range:	April 1997 – December 2017
Data source:	HES
Generate by:	UK Colorectal Cancer Intelligence Hub
Associated variables:	Primary procedure - PROCTYPE

Rules for derivation

The variable PROCNAME identifies the type of major resection for individuals who underwent a major resection as their primary procedure. The procedures were grouped according to their OPCS4 codes.

Stoma opened during major surgical resection - STOMA

Data item:	Stoma opened at major resection
Field name:	STOMA
Format:	Number
Description of field content:	0 – No stoma opened
	1 – Stoma opened
Valid event date range:	April 1997 – December 2017
Data source:	HES
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

The variable STOMA identifies individuals who had a stoma created during their major surgical resection for colorectal cancer. The creation of a stoma was identified using OPCS4 codes, either for a major surgical resection during which a stoma is created as standard, or where an OPCS code for a stoma was recorded on the same date as an alternative major surgical resection. If neither criteria was met it is assumed that no stoma was created at the time of major surgical resection.

Stoma open at 18 months following creation – STOMA18

Data item:	Stoma present at 18 months
Field name:	STOMA18
Format:	Number
Description of field content:	0 – Closed
	1 – Open
Valid event date range:	April 1997 – December 2017
Data source:	HES
Generate by:	UK Colorectal Cancer Intelligence Hub
Associated variables:	Stoma opened during major surgical resection - STOMA

Rules for derivation

The variable STOMA18 identifies, for individuals who had a stoma created during their major surgical resection for colorectal cancer, whether the stoma remained open at 18 months from creation. The reversal of a stoma was indicated by the presence of specified OPCS4 codes during a spell in hospital within 18 months of creation of the stoma.

Individuals who had an abdominoperineal excision or pelvic exenteration as their initial major surgical resection could not have their stoma reversed and so were classified as open at 18 months.

Approach to major surgical resection - LAP

Data item:	Approach to major resection
Field name:	LAP
Format:	Number
Description of field content:	0 – Open
	1 – Laparoscopic
	2 – Converted from laparoscopic to open
Valid event date range:	April 2006 – December 2017
Data source:	HES
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable identified major surgical resections which were undertaken laparoscopically or converted. These procedures were identified where an additional, relevant, OPCS4 code specifying laparoscopic surgery was recorded on the same date as a major surgical resection. Any major surgical resection without a specified OPCS4 code was assumed to be open.

Data quality

Codes for laparoscopic surgery were not introduced into the OPCS4 system until April 2006 and so analyses investigating use of laparoscopic surgery prior to this date are not feasible with this data source.

Procedures converted from laparoscopic to open - CONVERT

Definition

Data item:	Laparoscopic conversions
Field name:	CONVERT
Format:	Number
Description of field content:	0 – Not converted
	1 – Converted
	99 – Not laparoscopic
Valid event date range:	April 2006 – December 2017
Data source:	HES
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable identifies laparoscopic procedures which were attempted laparoscopically but were subsequently converted to an open operation. These procedures were identified where an

V1.0 September 2020

additional, relevant OPCS4 code specifying converted or successful laparoscopic surgery was recorded on the same date as a major surgical resection. Any major surgical resection without a specified OPCS4 code was assumed to be not laparoscopic (open).

Data quality

Codes for laparoscopic surgery were not introduced into the OPCS4 system until April 2006 and so analyses investigating use of laparoscopic surgery prior to this date are not feasible with this data source.

Type of laparoscopic procedure - APPTYPE

Data item:	Robotic laparoscopic procedures
Field name:	APPTYPE
Format:	Number
Description of field content:	0 – Standard laparoscopic
	1 – Robotic laparoscopic
	99 – Not laparoscopic
Valid event date range:	April 2006 – December 2017
Data source:	HES
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

This variable identifies major surgical resections which were performed as a robotic laparoscopic procedure. These procedures were identified where an additional, relevant, OPCS4 code specifying robotic laparoscopic surgery was recorded on the same date as a major surgical resection. Any major surgical resection without a specified OPCS4 code was assumed to be open.

Data quality

Codes for laparoscopic surgery were not introduced into the OPCS4 system before April 2006 and so analyses investigating use of laparoscopic surgery prior to this data are not feasible with this data source.

ACPGBI membership status of the overseeing consultant - ACPGBI

Data item:	ACPGBI membership status
Field name:	ACPGBI
Format:	Number
Description of field content:	0 – Non-member of the ACPGBI 1 – Member of the ACPGBI
Valid event date range:	December 2012-December 2016
Data source:	Association of Coloproctology of Great Britain and Ireland Clinical Outcomes Publication
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

The variable ACPGBI flags individuals recorded as a member of the Association of Coloproctology of Great Britain and Ireland who were identified through the Clinical Outcomes Publication (https://www.acpgbi.org.uk/surgeon-outcomes/). The surgeons were identified using the GMC number recorded alongside the procedure in question in the HES data.

Data quality

As the clinical outcomes publication was introduced in 2012 this variable is only covering procedures performed between December 2012 and December 2016.

ONCOLOGY

Neoadjuvant rectal cancer treatment - RECTALRT

Data item:	Radiotherapy category
Field name:	RECTALRT
Format:	Number
Description of field content:	0 – No radiotherapy
	1 – Short course radiotherapy with immediate surgery
	2 – Short course radiotherapy with delayed surgery
	3 – Long course chemoradiotherapy
	4 – Post-operative radiotherapy
	5 – Other radiotherapy
Valid event date range:	April 2009 – December 2017
Data source:	Linked NCRAS, HES and RTDS
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

For individuals who underwent a major surgical resection as the primary procedure for their rectal cancer, radiotherapy information was sought from the RTDS data. Radiotherapy use was grouped into six categories depending upon the dose given, how many times the patient attended for radiotherapy, and the interval between major surgical resection and the end of radiotherapy.

Number of attendances for radiotherapy (identified in the RTDS data)	Interval between end of radiotherapy treatment and major surgical resection	CORECT-R code	CORECT-R description
No record in RTDS data		0	**No radiotherapy**
4-5	35 days or less	1	**Short course radiotherapy with immediate surgery**
4-5	More than 35 days	2	**Short course radiotherapy with delayed surgery**
25, 28 or 30	365 days or less	3	**Long course radiotherapy**
25 attendances over multiple episodes where at least one has 10 or more attendances	365 days or less		
	Radiotherapy starting between 1 and 365 days after major resection	4	**Post-operative radiotherapy**
Any other number	365 days or less	5	**Other radiotherapy**

V1.0 September 2020

Adjuvant chemotherapy - ADJCT

Data item:	Adjuvant chemotherapy
Field name:	ADJCT
Format:	Number
Description of field content:	0 – No adjuvant chemotherapy
	1 – Adjuvant chemotherapy
Valid event date range:	January 2014 – December 2017
Data source:	Linked NCRAS, HES and SACT
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

For individuals who underwent a major surgical resection as the primary procedure for their colorectal cancer, chemotherapy information was sought from the SACT data. Adjuvant chemotherapy was defined as a chemotherapy regimen beginning within 6 months of their major surgical resection.

Data quality

This variable is only available for those diagnosed from 1st January 2014 onwards.

Type of adjuvant chemotherapy – ADJCT_TYPE

Data item:	Adjuvant chemotherapy type
Field name:	ADJCT_TYPE
Format:	Number
Description of field content:	0 – No adjuvant chemotherapy 1 – CAPOX 2 – FOLFOX 3 – CAP 4 – 5FU 5 - Other
Valid event date range:	January 2014 – December 2017
Data source:	Linked NCRAS, HES and SACT
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

For individuals who underwent a major surgical resection as the primary procedure for their colorectal cancer, chemotherapy information was sought from the SACT data. Adjuvant chemotherapy was defined as a chemotherapy regimen beginning within 6 months of their major surgical resection.

The chemotherapy type for those identified as having received adjuvant chemotherapy is derived from the regimen name within SACT.

The first regimen (based on the interval between major surgical resection and start date) is kept.

The following hierarchy is applied to multiple regimens starting at the same time;

1. Duplicates are excluded
2. Abbreviations of regimen names are replaced with full names (e.g. CAP, CAPE, CAPECIT replaced with CAPECITABINE)
3. If one regimen name is a substring of another the longer is kept. E.g. If FLUOR and FLUOROURACIL were both recorded the latter would be kept.
4. If monoclonal antibodies (mAbs) are recorded as a separate regimen these are included with the other regimen.
5. If fluorouracil is recorded in addition to a modified de Gramont regimen (MdG) then the MdG regimen is kept.
6. If a combination regimen and a component of that regimen are listed separately only the combination regimen is kept.
   o Any remaining multiples which contain any of the following are classed as combination chemotherapy; CAPECIT, OXALI, MDG, FLUOR, RALTIT, BEV, IRINO, AFLIB

The following are classed as adjuvant chemo:

CAPECITABINE

CAPECITABINE + OXALIPLATIN

CAPECITABINE + RT

CRC COMB

FLUOROURACIL

FLUOROURACIL + RT

OXALIPLATIN + MDG

Data quality

This variable is only available for those diagnosed from 1st January 2014 onwards.

Patient Reported Outcomes (PROMs)

The Patient Reported Outcomes (PROMs) data are available in isolation only. It is not possible to link any of the information in this dataset to any other tables or variables included in CORECT-R.

Patient and tumour characteristics

Age at colorectal cancer diagnosis - age

Data item:	Age
Field name:	age
Format:	Number
Description of field content:	Age at the time of colorectal cancer diagnosis, in years
Data source:	NCRAS cancer registration data
Generate by:	UK Colorectal Cancer Intelligence Hub

Sex - sex

Data item:	Sex
Field name:	sex
Format:	Number
Description of field content:	1 – Male 2 - Female
Data source:	NCRAS cancer registration data
Generate by:	UK Colorectal Cancer Intelligence Hub

Cancer site- site

Data item:	Cancer site
Field name:	site
Format:	Number
Description of field content:	1 – Colon 2 – Rectosigmoid 3 – Rectum
Data source:	NCRAS cancer registration data
Generate by:	UK Colorectal Cancer Intelligence Hub

Dukes stage- stage

Data item:	Dukes stage
Field name:	stage
Format:	Number
Description of field content:	1 – A
	2 - B
	3 – C
	4 – D
	9 - Unknown
Data source:	NCRAS cancer registration data
Generate by:	UK Colorectal Cancer Intelligence Hub

Diagnosis year - diagyear

Data item:	Colorectal cancer diagnosis date
Field name:	diagyear
Format:	Date
Description of field content:	2010
	2011
Data source:	NCRAS cancer registration data
Generate by:	UK Colorectal Cancer Intelligence Hub

IMD quintile – IMD2007

Data item:	Socioeconomic status
Field name:	IMD2007
Format:	Number
Description of field content:	1 – least deprived
	2- 2
	3 - 3
	4 - 4
	5 – most deprived
Data source:	NCRAS cancer registration data
Generate by:	UK Colorectal Cancer Intelligence Hub

Survey results

YOUR HEALTH TODAY

What treatments have you received for your colorectal cancer – Q1_1

Data item:	Radiotherapy
Field name:	Q1_1
Format:	Number
Description of field content:	1 – Radiotherapy
	2 – No radiotherapy

What treatments have you received for your colorectal cancer – Q1_2

Data item:	Chemotherapy
Field name:	Q1_2
Format:	Number
Description of field content:	1 – Chemotherapy
	2 – No chemotherapy

What treatments have you received for your colorectal cancer – Q1_3

Data item:	Surgery
Field name:	Q1_3
Format:	Number
Description of field content:	1 – Surgery
	2 – No surgery

How long is it since you completed your initial treatment for colorectal cancer? – Q2

Data item:	Time since treatment
Field name:	Q2
Format:	Number
Description of field content:	1 – I am still having my initial treatment
	2 – It is less than 3 months since my initial treatment
	3 – It is between 3 and 12 months since my initial treatment
	4 – It is between 1 and 5 years since my initial treatment
	5 – It is more than 5 years since my initial treatment
	6 – Don't know/can't remember

How has your colorectal cancer responded to treatment? – Q3

Data item:	Response to treatment
Field name:	Q3
Format:	Number
Description of field content:	1 – My colorectal has responded fully to treatment (I am in remission)
	2 – My colorectal cancer has been treated but is still present
	3 – My colorectal cancer has not been treated at all
	4 – My colorectal cancer has come back after it was originally treated
	5 – I am not certain what is happening with my colorectal cancer

If you have a stoma (e.g. colostomy) is it: – Q4

Data item:	Stoma
Field name:	Q4
Format:	Number
Description of field content:	1 – Still present
	2 – Reversed
	3 – This does not apply to me

Mobility – Q5

Data item:	Mobility
Field name:	Q5
Format:	Number
Description of field content:	1 – I have no problems in walking about
	2 – I have slight problems in walking about
	3 - I have moderate problems in walking about
	4 – I have severe problems in walking about
	5 – I am unable to walk about

Self-care – Q6

Data item:	Self-care
Field name:	Q6
Format:	Number
Description of field content:	1 – I have no problems washing or dressing myself
	2 – I have slight problems washing or dressing myself
	3 - I have moderate problems washing or dressing myself
	4 – I have severe problems washing or dressing myself
	5 – I am unable to wash or dress myself

Usual activities – Q7

Data item:	Usual activities
Field name:	Q7
Format:	Number
Description of field content:	1 – I have no problems doing my usual activities
	2 – I have slight problems doing my usual activities
	3 - I have moderate problems doing my usual activities
	4 – I have severe problems doing my usual activities
	5 – I am unable to do my usual activities

Pain/discomfort – Q8

Data item:	Pain/discomfort
Field name:	Q8
Format:	Number
Description of field content:	1 – I have no pain or discomfort
	2 – I have slight pain or discomfort
	3 - I have moderate pain or discomfort
	4 – I have severe pain or discomfort
	5 – I have extreme pain or discomfort

Anxiety/depression – Q9

Data item:	Anxiety/depression
Field name:	Q9
Format:	Number
Description of field content:	1 – I am not anxious or depressed 2 – I am slightly anxious or depressed 3 - I am moderately anxious or depressed 4 – I am severely anxious or depressed 5 – I am extremely anxious or depressed

YOUR HEALTH AND WELLBEING IN THE PAST WEEK

I have swelling or cramps in my stomach area – Q10

Data item:	Swelling or cramps
Field name:	Q10
Format:	Number
Description of field content:	1 – Not at all 2 – A little bit 3 – Somewhat 4 – Quite a bit 5 – Very much

I am losing weight – Q11

Data item:	Losing weight
Field name:	Q11
Format:	Number
Description of field content:	1 – Not at all 2 – A little bit 3 – Somewhat 4 – Quite a bit 5 – Very much

I have control of my bowels – Q12

Data item:	Control of bowels
Field name:	Q12
Format:	Number
Description of field content:	1 – Not at all
	2 – A little bit
	3 – Somewhat
	4 – Quite a bit
	5 – Very much

I can digest my food well – Q13

Data item:	Digest food
Field name:	Q13
Format:	Number
Description of field content:	1 – Not at all
	2 – A little bit
	3 – Somewhat
	4 – Quite a bit
	5 – Very much

I have diarrhoea – Q14

Data item:	Diarrhoea
Field name:	Q14
Format:	Number
Description of field content:	1 – Not at all
	2 – A little bit
	3 – Somewhat
	4 – Quite a bit
	5 – Very much

I have a good appetite – Q15

Data item:	Good appetite
Field name:	Q15
Format:	Number
Description of field content:	1 – Not at all
	2 – A little bit
	3 – Somewhat
	4 – Quite a bit
	5 – Very much

I like the appearance of my body – Q16

Data item:	Appearance
Field name:	Q16
Format:	Number
Description of field content:	1 – Not at all
	2 – A little bit
	3 – Somewhat
	4 – Quite a bit
	5 – Very much

I have difficulty urinating – Q17

Data item:	Difficulty urinating
Field name:	Q17
Format:	Number
Description of field content:	1 – Not at all
	2 – A little bit
	3 – Somewhat
	4 – Quite a bit
	5 – Very much

I urinate more frequently than usual – Q18

Data item:	Frequency of urination
Field name:	Q18
Format:	Number
Description of field content:	1 – Not at all 2 – A little bit 3 – Somewhat 4 – Quite a bit 5 – Very much

I leak urine – Q19

Data item:	Leaking urine
Field name:	Q19
Format:	Number
Description of field content:	1 – Not at all 2 – A little bit 3 – Somewhat 4 – Quite a bit 5 – Very much

Do you have an ostomy appliance/stoma? – Q20

Data item:	Ostomy/stoma
Field name:	Q20
Format:	Number
Description of field content:	1 – No 2 – Yes

I am embarrassed by my ostomy appliance/stoma – Q21

Data item:	Embarrassed
Field name:	Q21
Format:	Number
Description of field content:	1 – Not at all 2 – A little bit 3 – Somewhat 4 – Quite a bit 5 – Very much

Caring for my ostomy appliance/stoma is difficult – Q22

Data item:	Care is difficult
Field name:	Q22
Format:	Number
Description of field content:	1 – Not at all
	2 – A little bit
	3 – Somewhat
	4 – Quite a bit
	5 – Very much

Do you have any difficulty in controlling your bowels (e.g. any accidents)? – Q23

Data item:	Difficulty controlling bowels
Field name:	Q23
Format:	Number
Description of field content:	1 – No
	2 – Yes

If yes (to Q23) how often do you have difficulties? – Q24

Data item:	How often do you have difficulties
Field name:	Q24
Format:	Number
Description of field content:	1 – Monthly
	2 – Weekly
	3 – Daily
	4 – Constantly
	5 – It varies

In the past week, on how many days have you done a total of 30 minutes or more of physical activity which was brisk enough to raise your heartbeat? – Q25

Data item:	Physical activity
Field name:	Q25
Format:	Number
Description of field content:	1 – None
	2 – 1 day
	3 – 2 days
	4 – 3 days
	5 – 4 days
	6 – 5 days
	7 – 6 days
	8 – 7 days

YOUR HEALTH AND WELLBEING IN THE PAST MONTH

Have you had any difficulty in maintaining your independence? – Q26

Data item:	Independence
Field name:	Q26
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty in carrying out your domestic chores? (E.g. cleaning, gardening, cooking, shopping) – Q27

Data item:	Domestic chores
Field name:	Q27
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty with managing your own personal care? (E.g. bathing, dressing, washing) – Q28

Data item:	Personal care
Field name:	Q28
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty with looking after those who depend on you? (E.g. children, dependent adults, pets) – Q29

Data item:	Looking after those who depend on you
Field name:	Q29
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have any of those close to you (e.g. partner, children, parents) had any difficulty with the support available to them? – Q30

Data item:	Difficulty with support
Field name:	Q30
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty with benefits? (E.g. statutory sick pay, attendance allowance, disability living allowance) – Q31

Data item:	Benefits
Field name:	Q31
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any financial difficulties? – Q32

Data item:	Financial difficulties
Field name:	Q32
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty with financial services? (E.g. loans, mortgages, pensions, insurance) – Q33

Data item:	Financial services
Field name:	Q33
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty concerning your work? (or education if you are a student) – Q34

Data item:	Difficulty concerning work
Field name:	Q34
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty in planning for your own or your family future? (E.g. care of dependents, legal issues, business affairs) – Q35

Data item:	Planning for the future
Field name:	Q35
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty with communicating with those closest to you? (E.g. partner, children, parents) – Q36

Data item:	Communication
Field name:	Q36
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty with communicating with others? (E.g. friends, neighbours, colleagues, dates) – Q37

Data item:	Communicating with others
Field name:	Q37
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty concerning sexual matters? – Q38

Data item:	Sexual matters
Field name:	Q38
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty concerning plans to have a family? – Q39

Data item:	Plans to have a family
Field name:	Q39
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty concerning your appearance or body image? – Q40

Data item:	Appearance/body image
Field name:	Q40
Format:	Number
Description of field content:	1 – No difficulty 2 – A little 3 – Quite a bit 4 – Very much 5 – Does not apply

Have you felt isolated? – Q41

Data item:	Isolation
Field name:	Q41
Format:	Number
Description of field content:	1 – No difficulty 2 – A little 3 – Quite a bit 4 – Very much 5 – Does not apply

Have you had any difficulty with getting around? (E.g. transport, car parking, your mobility) – Q42

Data item:	Getting around
Field name:	Q42
Format:	Number
Description of field content:	1 – No difficulty 2 – A little 3 – Quite a bit 4 – Very much 5 – Does not apply

Have you had any difficulty with where you live? (E.g. space, access, damp, heating, neighbours, security) – Q43

Data item:	Difficulties with where you live
Field name:	Q43
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty in carrying out your recreational activities? (E.g. hobbies, pastimes, social pursuits) – Q44

Data item:	Recreational activities
Field name:	Q44
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty with your plans to travel or take a holiday? – Q45

Data item:	Travel and holidays
Field name:	Q45
Format:	Number
Description of field content:	1 – No difficulty
	2 – A little
	3 – Quite a bit
	4 – Very much
	5 – Does not apply

Have you had any difficulty with any other area of your life? – Q46

Data item:	Other aspects of life
Field name:	Q46
Format:	Number
Description of field content:	1 – No difficulty 2 – A little 3 – Quite a bit 4 – Very much 5 – Does not apply

I have fears about my cancer spreading – Q47

Data item:	Cancer spreading
Field name:	Q47
Format:	Number
Description of field content:	1 – Strongly agree 2 – Agree 3 – Neither agree nor disagree 4 – Disagree 5 – Strongly disagree 6 – Does not apply to me

I have fears about my cancer coming back – Q48

Data item:	Cancer returning
Field name:	Q48
Format:	Number
Description of field content:	1 – Strongly agree 2 – Agree 3 – Neither agree nor disagree 4 – Disagree 5 – Strongly disagree 6 – Does not apply to me

I have fears about death and dying – Q49

Data item:	Death and dying
Field name:	Q49
Format:	Number
Description of field content:	1 – Strongly agree
	2 – Agree
	3 – Neither agree nor disagree
	4 – Disagree
	5 – Strongly disagree

I experience memory loss – Q50

Data item:	Memory loss
Field name:	Q50
Format:	Number
Description of field content:	1 – Strongly agree
	2 – Agree
	3 – Neither agree nor disagree
	4 – Disagree
	5 – Strongly disagree

I have trouble sleeping – Q51

Data item:	Trouble sleeping
Field name:	Q51
Format:	Number
Description of field content:	1 – Strongly agree
	2 – Agree
	3 – Neither agree nor disagree
	4 – Disagree
	5 – Strongly disagree

I have trouble concentrating – Q52

Data item:	Trouble concentrating
Field name:	Q52
Format:	Number
Description of field content:	1 – Strongly agree 2 – Agree 3 – Neither agree nor disagree 4 – Disagree 5 – Strongly disagree

I always feel tired – Q53

Data item:	Always feel tired
Field name:	Q53
Format:	Number
Description of field content:	1 – Strongly agree 2 – Agree 3 – Neither agree nor disagree 4 – Disagree 5 – Strongly disagree

I experience mood swings – Q54

Data item:	Mood swings
Field name:	Q54
Format:	Number
Description of field content:	1 – Strongly agree 2 – Agree 3 – Neither agree nor disagree 4 – Disagree 5 – Strongly disagree

I am often irritable – Q55

Data item:	Often irritable
Field name:	Q55
Format:	Number
Description of field content:	1 – Strongly agree
	2 – Agree
	3 – Neither agree nor disagree
	4 – Disagree
	5 – Strongly disagree

OVERALL SUPPORT AND CARE

Do you have an up-to-date written care plan? – Q56

Data item:	Care plan
Field name:	Q56
Format:	Number
Description of field content:	1 – Yes, definitely
	2 – Yes, I think so
	3 – No
	4 – I do not need a care plan
	5 – Don't know

Do you have a named nurse you can contact if you have a worry about your cancer care? – Q57

Data item:	Named nurse
Field name:	Q57
Format:	Number
Description of field content:	1 – Yes
	2 – No
	3 – Don't know

Do you know who to contact if you have a concern about any aspect of living with or after cancer? – Q58

Data item:	Contact
Field name:	Q58
Format:	Number
Description of field content:	1 – Yes, definitely
	2 – Yes, I think so
	3 - No

Do you think that hospital staff did everything they could to support you following your cancer treatment? – Q59

Data item:	Support following cancer treatment
Field name:	Q59
Format:	Number
Description of field content:	1 – Yes, all of the time
	2 – Only some of the time
	3 – Never
	4 – I did not need support

Do you think that GPs and nurses at your general practice do everything they can to support you following your cancer treatment? – Q60

Data item:	Support following cancer treatment - GP
Field name:	Q60
Format:	Number
Description of field content:	1 – Yes, all of the time
	2 – Only some of the time
	3 – Never
	4 – My general practice is not involved
	5 – I do not need support

Following your initial cancer treatment have you been given enough care and help from health and social services – Q61

Data item:	Support following cancer treatment – health and social services
Field name:	Q61
Format:	Number
Description of field content:	1 – Yes, all of the time
	2 – Yes, to some extent
	3 – No
	4 – I did not need help from health or social services
	5 – Don't know/can't remember

Do you consider yourself to be a smoker? – Q62

Data item:	Smoker
Field name:	Q62
Format:	Number
Description of field content:	1 – Smoker
	2 – Ex-smoker
	3 – Non-smoker

If an ex-smoker, how long ago did you stop? – Q63

Data item:	Time since smoking
Field name:	Q63
Format:	Number
Description of field content:	1 – Less than 1 year
	2 – Less than 2 years
	3 – Less than 3 years
	4 – Less than 4 years
	5 – Less than 5 years
	6 – More than 5 years

Did you receive any advice or information on any of the following issues? – Q64_1

Data item:	Diet and lifestyle
Field name:	Q64_1
Format:	Number
Description of field content:	1 – Yes
	0 – No

Did you receive any advice or information on any of the following issues? – Q64_2

Data item:	Physical activity and exercise
Field name:	Q64_2
Format:	Number
Description of field content:	1 – Yes
	0 – No

Did you receive any advice or information on any of the following issues? – Q64_3

Data item:	Financial help or benefits
Field name:	Q64_3
Format:	Number
Description of field content:	1 – Yes
	0 – No

Did you receive any advice or information on any of the following issues? – Q64_4

Data item:	Free prescriptions
Field name:	Q64_4
Format:	Number
Description of field content:	1 – Yes
	0 – No

Did you receive any advice or information on any of the following issues? – Q64_5

Data item:	Returning to or staying in work
Field name:	Q64_5
Format:	Number
Description of field content:	1 – Yes
	0 – No

Did you receive any advice or information on any of the following issues? – Q64_6

Data item:	Information/ advice for family/ friends or carers
Field name:	Q64_6
Format:	Number
Description of field content:	1 – Yes
	0 – No

Did you receive any advice or information on any of the following issues? – Q64_7

Data item:	Physical aspects of living with and beyond cancer
Field name:	Q64_7
Format:	Number
Description of field content:	1 – Yes
	0 – No

Did you receive any advice or information on any of the following issues? – Q64_8

Data item:	Psychological aspects of living with and after cancer
Field name:	Q64_8
Format:	Number
Description of field content:	1 – Yes 0 – No

Did you receive any advice or information on any of the following issues? – Q64_9

Data item:	How to access support groups
Field name:	Q64_9
Format:	Number
Description of field content:	1 – Yes 0 – No

Did you receive any advice or information on any of the following issues? – Q64_10

Data item:	I have all the information and advice I need
Field name:	Q64_10
Format:	Number
Description of field content:	1 – Yes 0 – No

Did you receive any advice or information on any of the following issues? – Q64_11

Data item:	I was not offered any of the above
Field name:	Q64_11
Format:	Number
Description of field content:	1 – Yes 0 – No

Would it have been helpful to have had more advice or information on any of the following issues – Q65_1

Data item:	Diet and lifestyle
Field name:	Q65_1
Format:	Number
Description of field content:	1 – Yes 0 – No

Would it have been helpful to have had more advice or information on any of the following issues – Q65_2

Data item:	Physical activity and exercise
Field name:	Q65_2
Format:	Number
Description of field content:	1 – Yes
	0 – No

Would it have been helpful to have had more advice or information on any of the following issues – Q65_3

Data item:	Financial help or benefits
Field name:	Q65_3
Format:	Number
Description of field content:	1 – Yes
	0 – No

Would it have been helpful to have had more advice or information on any of the following issues – Q65_4

Data item:	Free prescriptions
Field name:	Q65_4
Format:	Number
Description of field content:	1 – Yes
	0 – No

Would it have been helpful to have had more advice or information on any of the following issues – Q65_5

Data item:	Returning to or staying in work
Field name:	Q65_5
Format:	Number
Description of field content:	1 – Yes
	0 – No

Would it have been helpful to have had more advice or information on any of the following issues – Q65_6

Data item:	Information/ advice for family/ friends or carers
Field name:	Q65_6
Format:	Number
Description of field content:	1 – Yes
	0 – No

Would it have been helpful to have had more advice or information on any of the following issues – Q65_7

Data item:	Physical aspects of living with and beyond cancer
Field name:	Q65_7
Format:	Number
Description of field content:	1 – Yes
	0 – No

Would it have been helpful to have had more advice or information on any of the following issues – Q65_8

Data item:	Psychological aspects of living with and after cancer
Field name:	Q65_8
Format:	Number
Description of field content:	1 – Yes
	0 – No

Would it have been helpful to have had more advice or information on any of the following issues – Q65_9

Data item:	How to access support groups
Field name:	Q65_9
Format:	Number
Description of field content:	1 – Yes
	0 – No

Would it have been helpful to have had more advice or information on any of the following issues – Q65_10

Data item:	I have all the information and advice I need
Field name:	Q65_10
Format:	Number
Description of field content:	1 – Yes
	0 – No

ABOUT YOU

What year were you born? – Q66

Data item:	Year of birth
Field name:	Q66
Format:	Number
Description of field content:	Four digit year of birth

Are you male or female? – Q67

Data item:	Sex
Field name:	Q67
Format:	Number
Description of field content:	1 – Male
	2 - Female

Do you look after, or give any help or support (not as part of your paid employment) to family members, friends, neighbours or others because of either; - Long term physical or mental health/disability or – Problems relating to old age? – Q68

Data item:	Provision of support to others
Field name:	Q68
Format:	Number
Description of field content:	1 – Yes
	2 - No

Which of the following best describes your sexual orientation? – Q69

Data item:	Sexual orientation
Field name:	Q69
Format:	Number
Description of field content:	1 – Heterosexual/straight (opposite sex)
	2 – Bisexual (both sexes)
	3 – Gay or lesbian (same sex)
	4 – Other
	5 – Prefer not to say

Which statement best describes your living arrangements? – Q70

Data item:	Living arrangements
Field name:	Q70
Format:	Number
Description of field content:	1 – I live with partner/spouse/family/friends
	2 – I live alone
	3 – I live in a nursing home, hospital or other long term care home
	4 - Other

Do you have a long standing health condition? (Please include anything other than your cancer that has troubled you over a period of time or that could affect you over a period of time)– Q71

Data item:	Long term conditions
Field name:	Q71
Format:	Number
Description of field content:	1 – Yes
	2 – No
	3 – Don't know/can't say

Which, if any, of the following conditions do you have? – Q72_1

Data item:	Alzheimer's disease or dementia
Field name:	Q72_1
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_2

Data item:	Angina
Field name:	Q72_2
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_3

Data item:	Arthritis
Field name:	Q72_3
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_4

Data item:	Asthma or other chronic chest problem
Field name:	Q72_4
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_5

Data item:	Blindness or visual impairment
Field name:	Q72_5
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_6

Data item:	Deafness or hearing impairment
Field name:	Q72_6
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_7

Data item:	Diabetes
Field name:	Q72_7
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_8

Data item:	Epilepsy
Field name:	Q72_8
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_9

Data item:	Heart condition
Field name:	Q72_9
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_10

Data item:	High blood pressure
Field name:	Q72_10
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_11

Data item:	Kidney disease
Field name:	Q72_11
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_12

Data item:	Learning difficulty
Field name:	Q72_12
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_13

Data item:	Liver disease
Field name:	Q72_13
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_14

Data item:	Long term back problems
Field name:	Q72_14
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_15

Data item:	Long standing mental health problem
Field name:	Q72_15
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_16

Data item:	Long standing neurological problem
Field name:	Q72_16
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_17

Data item:	Another long standing condition
Field name:	Q72_17
Format:	Number
Description of field content:	1 – Yes
	0 – No

Which, if any, of the following conditions do you have? – Q72_18

Data item:	No long standing condition
Field name:	Q72_18
Format:	Number
Description of field content:	1 – Yes
	0 – No

What was your employment status before you were diagnosed with cancer – Q73

Data item:	Employment status – before cancer
Field name:	Q73
Format:	Number
Description of field content:	1 – Full time employment
	2 – Part time employment
	3 – Homemaker
	4 – Student (in education)
	5 – Retired
	6 – Unemployed and seeking work
	7 – Unemployed and unable to work for health reasons
	8 - Other

What was your employment status before you were diagnosed with cancer – Q74

Data item:	Employment status – current
Field name:	Q74
Format:	Number
Description of field content:	1 – Full time employment
	2 – Part time employment
	3 – Homemaker
	4 – Student (in education)
	5 – Retired
	6 – Unemployed and seeking work
	7 – Unemployed and unable to work for health reasons
	8 - Other

If you are currently employed at the moment, are you – Q75

Data item:	Working
Field name:	Q75
Format:	Number
Description of field content:	1 – Not working at all
	2 – Working less hours than usual
	3 – Working your usual hours
	4 – Working more hours than usual
	5 – This question does not apply to me

To which of these ethnic groups would you say you belong? – Q76

Data item:	Ethnicity
Field name:	Q76
Format:	Number
Description of field content:	1 – White British
	2 – White Irish
	3 – Any other white background
	4 – White and Black Caribbean
	5 – White and Black African
	6 – White and Asian
	7 – Any other Mixed background
	8 – Indian
	9 – Pakistani
	10 – Bangladeshi
	11 – Any other Asian background
	12 – Caribbean
	13 – African
	14 - Any other Black background
	15 – Chinese
	16 – Any other ethnic group

National Anal Cancer Dataset (variables included in the anal table only)

TUMOUR INFORMATION

Morphology group for anal cancers – USEMORPH_ANAL

Data item:	Morphology group of anal tumours
Field name:	USEMORPH_ANAL
Format:	Number
Description of field content:	1 – Squamous cell carcinoma
	2 – Basaloid and cloacogenic carcinoma
	3 – Adenocarcinoma
	4 – Other specified carcinoma
	5 – Unspecified carcinomas
	6 – Melanoma
	7 – Other specified cancer
	9 – Unspecified cancer
Valid event date range:	April 1997 to December 2017
Data source:	NCRAS data
Generated by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

The variable USEMORPH_ANAL is used classify the morphology of anal tumours. The classification is based on the ICD02 code for the tumour recorded within the cancer registry data.

Morphologies which don't fall into categories 1-7 are classified as category 9 – unspecified cancer.

Data quality

This variable includes only anal tumours (C21). Morphology groupings for tumours of the colon, rectosigmoid or rectum can be found in USEMORPH.

Stage of anal cancer – USESTAGE_ANAL

Data item:	Stage of anal cancer
Field name:	USESTAGE_ANAL
Format:	Number
Description of field content:	0 – High grade AIN/ stage 0 1 – stage 1 2 – stage 2a 3 – stage 2b 4 – stage 3a 5 – stage 3b 6 – stage 3c 7 – stage 4 9 - unknown
Valid event date range:	April 1997 – December 2017
Data source:	NCRAS
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

The USESTAGE_ANAL variable classifies the stage of anal cancer using multiple sources within the NCRAS data to provide a single source of information.

Tumours are classified using the STAGEB field from NCRAS in the first instance.

AIN stage	T stage	N stage	M stage	CORECT-R code	CORECT-R description
3				0	**High grade AIN/0**
	1	0	0	1	**1**
	2	0	0	2	**2a**
	3	0	0	3	**2b**
	1	1	0	4	**3a**
	2	1	0		
	4	0	0	5	**3b**
	3	1	0	6	**3c**
	4	1	0		
	Any	Any	1	7	**4**
				9	**Unknown**

Abdominoperineal resection – APE

Data item:	Abdominoperineal resection
Field name:	APE
Format:	Number
Description of field content:	0 – No APE 1 – APE
Valid event date range:	April 1997 – December 2017
Data source:	NCRAS
Generate by:	UK Colorectal Cancer Intelligence Hub

Rules for derivation

The APE variable identified individuals with anal cancer who underwent an abdominoperineal resection of their anal tumour. APE resections are identified using OPCS 4 codes.

OPCS code	Description	CORECT-R code	CORECT-R description
H331	Abdominoperineal excision of rectum and end colostomy	1	**Abdominoperineal resection**
H337	Perineal resection of rectum HFQ		

Datasets awaiting linkage

As outlined at the beginning of this catalogue, CORECT-R will eventually incorporate other UK CRC data sets. Details of those currently awaiting linkage which cover the English NHS can be found in this section.

[Cancer Waiting Times](#)

[Diagnostic Imaging Dataset (DID)](#)

[National Cancer Diagnosis Audit](#)

This section also includes details of datasets which cover the Scottish NHS. Currently, permissions are in place for a linked dataset of Scottish CRC patients for the period 2006-2019. The Public Benefit Privacy Panel for Health and Social Care approved this study in October 2018 (Project number: 1718-0026). More details are available on the Edinburgh Health Economics website here: https://blogs.ed.ac.uk/ectu_ehe/research/bciuk/

At present, storage and access to this data is limited to the named researchers on the project within Scotland and within the agreed time-frame for the storage of the data (currently permissions are in place until May 2022). However, permissions are being sought to indefinitely extend the lifecycle of the project and store the Scottish data within the trusted research environment alongside the other data sets included in CORECT-R. In what follows is a description of some of the key CRC Scottish datasets alongside an overview of the variables included in them. All data sets are ready to be linked to one another through a unique patient ID.

[Scottish Cancer Registry (SMR06)](#)

[National Records of Scotland (NRS) Deaths](#)

[Scottish Morbidity Records (SMR)](#)

[Chemocare](#)

[Quality Performance Indicators (QPI)](#)

Cancer Waiting Times

The National Cancer Waiting Times (CWT) Monitoring Data Set is gathered by NHS England and it supports the continued management and monitoring of the following waiting times relevant to colorectal cancer: These are

- A maximum two week wait from an urgent GP referral for suspected cancer to DATE FIRST SEEN by a specialist
- A maximum one month (31-day) wait from diagnosis (CANCER TREATMENT PERIOD START DATE) to First Definitive Treatment
- A maximum two month (62-day) wait from urgent GP referral for suspected cancer to First Definitive Treatment
- A maximum 62-day wait from referral from a cancer screening programme to first treatment
- A maximum 62-day wait from a CONSULTANTS decision to upgrade the urgency of a PATIENT they suspect to have cancer to first treatment
- A maximum 31-day wait for all subsequent treatments for new cases of primary and recurrent cancer where an Anti-Cancer Drug Regimen, surgery or Radiotherapy is the chosen CANCER TREATMENT MODALITY

The CWT information in CORECT-R is limited to only those records available for individuals who have had a diagnosis of colorectal and anal cancers recorded by the NCRAS.

Data Summary

Data Provider	**NHS Digital**
Temporality of the data	*1st January 2009-31st December 2015*
Geographical extent	England
Further information	https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/cancerwaitingtimescwt
Data dictionary	https://www.datadictionary.nhs.uk/data_dictionary/messages/clinical_data_sets/data_sets/national_cancer_waiting_times_monitoring_data_set_fr.asp

Diagnostic Imaging Dataset (DID)

The Diagnostic Imaging Dataset (DID) is a central collection of detailed information about diagnostic imaging tests carried out on NHS patients, extracted from local radiology information systems. The DID captures information about referral source, details of the test (type of test and body site), demographic information such as GP registered practice, patient postcode, ethnicity, gender and date of birth, plus data items about different events (date of imaging request, date of imaging, date of reporting, which allows calculation of time intervals.

The DID information in CORECT-R is limited to only those records available for individuals who have had a diagnosis of colorectal and anal cancers recorded by the NCRAS. Currently linked data are only available for those diagnosed 1st April 2012- 31st December 2015.

Data Summary

Data Provider	**NHS Digital**
Temporality of the data	*1st April 2012 to 31st December 2015*
Geographical extent	England
Further information	https://www.england.nhs.uk/statistics/statistical-work-areas/diagnostic-imaging-dataset/
Data dictionary	https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/diagnostic-imaging-data-set

V1.0 September 2020

National Cancer Diagnosis Audit

The National Cancer Diagnosis Audit (NCDA) in England collected primary care data on the diagnostic pathway for 17,042 patients diagnosed with cancer in 2014.

The NCDA cohort was created by selecting all malignant cancer cases (excluding non-melanoma skin cancer) diagnosed in 2014, as registered by NCRAS. Participation on the NCDA was optional and verified GPs entered primary care information on patients registered at their practice whilst diagnosed with cancer. With the exception of dates, GPs selected pre-defined answers from a drop-down list. Data were not entered on patients whose cancer was detected through screening but the data has been retained in the dataset for completion.

The audit in England was conducted between Sept 2016 and Feb 2017 and 439 practices (about 5% of all practices) submitted data. Of the practices that participated, 365 practices entered data on ≥95% of patients from the audit list. The NCDA represents 6% of the total cancers registered in 2014 and the distribution of age and cancer types were similar to the total cohort of cancers registered in 2014. The participating practices were similar to non-participating practices, with respect to age, urban/rural location and practice -based patient experience measures, however practices patriating in the NCDA were somewhat larger and had slightly fewer patients per GP.

The NCDA information in CORECT-R is limited to only those records available for individuals who have had a diagnosis of colorectal and anal cancers recorded by the NCRAS.

Data Summary

Data Provider	**National Cancer Diagnosis Audit**
Temporality of the data	*2014*
Geographical extent	England
Further information	http://www.cancerresearchuk.org/health-professional/diagnosis/national-cancer-diagnosis-audit
Data dictionary	

Scottish Cancer Registry (SMR06)

This dataset includes information on all new diagnoses of cancer occurring within Scotland. These data are collected by Public Health Scotland and contain diagnostic, staging and treatment information on all cancers, including every colorectal and anal cancer diagnosed in Scotland. The SMR06 data is routinely linked with NRS deaths data and hospital admissions data as part of the Information Services Division linked data catalogue. The CORECT-R Scotland database contains the SMR06 records for all patients who had a diagnosis of colorectal cancer between January 2006 and December 2018. All SMR06 records for non-colorectal cancer diagnoses are included if the patient also had a non-colorectal cancer diagnosis during the study period.

Data Summary

Data Provider	**Public Health Scotland**
Temporality of the data	*January 2006 – December 2018*
Geographical extent	Scotland
Data tables	Patient
Further information and data dictionary	https://www.isdscotland.org/Health-Topics/Cancer/Scottish-Cancer-Registry/Cancer-Metadata/_docs/SMR06-Current-Dataset-from-20190101.pdf

National Records of Scotland (NRS) Deaths

The NRS are responsible for the registration of all life events occurring in Scotland including births, deaths, marriages, civil partnerships and adoptions. They are also responsible for Scottish census. The vital events deaths data set contains information on deaths including the leading cause of death, other causes of death, place of death, duration of illness and much more. For the purposes of the CORECT-R Scotland data, deaths data were collected for any CRC patient who died throughout the study period (2006-2019).

Data Summary

Data Provider	**National Records of Scotland**
Temporality of the data	*January 2006 – August 2019*
Geographical extent	Scotland
Data tables	NRS Vital Events- Deaths
Further information and data dictionary	https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/deaths
	https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/deaths/deaths-background-information

Scottish Morbidity Records (SMR)

The Scottish Morbidity Records (SMR) contain healthcare data for individual patients. There are four main SMR series for the general types of healthcare received during an episode and/or the nature or status of the patient. Those are outpatient attendance (SMR00), general/acute inpatient and day case (SMR01), maternity inpatient and day case (SMR02) and mental health inpatient and day case (SMR04).

The largest of the SMR series, the SMR01 database, contains episode level data for all general/acute inpatient or day cases in Scottish NHS hospitals or Scottish NHS beds in non-NHS-institutions. An SMR01 record is generated for an inpatient or day case for the following reasons: when they are admitted to an NHS hospital from a location external to the NHS; when they are admitted to a contracted NHS bed in a non-NHS-institution; when they change speciality; when they transfer from another NHS hospital; when they change consultant but not specialty. Further, an SMR01 record is generated when an inpatient moves into and/or out of one of the valid significant facilities and when they return to hospital after been on pass for more than 5 days.

The Scottish CORECT-R linked data set contains all concurrent and historic (from January 1997) SMR01 records for patients who had a colorectal cancer diagnosis during the study period (January 2006-December 2018) as identified from the SMR06 registry.

Data Summary

Data Provider	**Public Health Scotland**
Temporality of the data	*January 1997-August 2019*
Geographical extent	Scotland
Data tables	SMR01- General/acute inpatient and day case
Further information	https://www.ndc.scot.nhs.uk/Dictionary-A-Z/Definitions/index.asp?Search=S&ID=460&Title=SMR01%20-%20General/Acute%20Inpatient%20and%20Day%20Case
Data dictionaries	https://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets//SMR01-General-Acute-Inpatient-and-Day-Case/

Chemocare

Chemocare data is held separately by the three regional cancer networks in Scotland. Those are the South East Cancer Network (SCAN), West of Scotland Cancer Network (WoSCAN) and the North of Scotland Cancer Network (NoSCAN). Each network uses the Chemocare system to electronically record chemotherapy prescribing information for all cancer patients treated within their respective cancer network. This includes information on the patient including their height and weight, the drugs and doses prescribed, regimens etc.

The Chemocare information in the Scottish CORECT-R is limited to only those individuals who have had a diagnosis of colorectal or anal cancer between January 2013 and June 2019. These records can be linked to all other datasets held in the Scottish CORECT-R. Given that Chemocare is more up to date than the records held in SMR06, Chemocare may contain patients who are not present in the SMR06   registry. Moreover, there are some differences in the recording of data between the three cancer networks. Where this is the case, this is highlighted in the data table.

Data Summary

Data Provider	**WoSCAN; NoSCAN; SCAN**
Temporality of the data	*January 2013 - June 2019*
Geographical extent	Scotland
Data tables	Chemocare
Further information	
Data dictionary	There is currently no data dictionary for Chemocare

Quality Performance Indicators (QPI)

As part of the NHS Scotland Healthcare Quality Strategy in 2010, the National Quality Performance Indicators (QPIs) were developed. These are a set of cancer specific, outcome focussed, evidence based indicators used to drive quality improvement in cancer care. Currently, these indicators are in place for 18 tumour types, including colorectal cancer. The QPI audits are carried out by each of the three cancer networks across Scotland and the data are curated by Public Health Scotland.

The Scottish CORECT-R include the Colorectal QPI indicators for those patients who had a colorectal cancer diagnosis during the period January 2013 to March 2018.

Data Summary

Data Provider	**Public health Scotland (collected by WoSCAN; NoSCAN; SCAN)**
Temporality of the data	*January 2013 – March 2019*
Geographical extent	Scotland
Further information	https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?ID=2&SubID=21
Data dictionary	https://www.isdscotland.org/Health-Topics/Cancer/Cancer-Audit/docs/Colorectal/Colorectal-Cancer-QPI-Dataset-V3-4-Final.pdf

Data profile: Scottish Colorectal Cancer Dataset

Content

In what follows, we outline four data tables from the first batch of the CORECT-R Scotland data. Those are for NRS deaths, SMR06, SMR01, QPI and Chemocare. Some variables within the datasets are derived using pre-existing variables in that data set. The main spine of patients comes from the SMR06 registry. From here, all patients can be linked to their SMR01 records via their master ID. Further, if the patient died during the study period their SMR06 record can also be linked to their NRS deaths record. Patient records can also be linked to Chemocare and QPI, though it is possible that some patients who are present in Chemocare, are not present in SMR06, due to Chemocare covering a more recent time period than SMR06.

Data Summary

Data Provider	National Safe Haven Scotland (Project number: 1718-0026)
Temporality of the data	*January 1997-June 2019 (dates vary depending on data table)*
Geographical extent	Scotland
Data tables	SMR06
	NRS Deaths
	SMR01
	QPI
	Chemocare

SMR06 data set (2006-2018)				
Data Item	Variable Name	Description of field content	Format	Further info
Identifiers				
Master ID	master_index	Pseudononymised person ID	Text	
Patient information				
Date of birth	dob_fmt	Date of Birth: MMYYYY	Date	
Gender	Sex	Sex 0 = Not Known; 1 = Male; 2 = Female; 9 = (includes not stated by patient, or not recorded)	Number	
Age	age_in_years	Age at incidence date	Number	Derived from dob_fmt
10 year age bands	age_bands	"<35" "35-44" "45-54" "55-64" "65-74" "75-84" "85+"	Number	Derived from age_in_years
Date of death	dod_fmt	Date of Death: DDMMYYYY	Date	
Vital status	died	Vital status of patient. 0 = No, 1 = Yes	Number	Derived
Age of death	age_died	Age at death	Number	
Cause of death	cause_i	Cause of death i where i = 1,…,8. ICD10 Codes.	Text	
Died from crc	crc_death	CRC on death certificate as cause of death. 1 = Yes, 0 = No	Number	
Emigration date	embarkation_date	The date of emigration from Scotland: DDMMYYYY	Date	
End of follow up	end_of_follow_up	Date last observed (died, left country or censored): DDMMYYYY	Date	Derived from date of death, left country or final patient death.
Left Scotland	left_scotland	Patient left Scotland (based on presence of embarkation date)	Number	Derived from embarkation date
Survival time	survival	Years and months from date of incidence to date of follow up	Number	Derived from date of incidence and date of follow up.
Incidence date	incidence_date_fmt	Date of incidence: DDMMYYYY	Date	
Death certificate	death_certificate	The case has FIRST come to light ONLY as a result of a death. 0 = No, 1 = Yes	Number	
Death certificate only	death_certificate_only	The case has been registered from the death certificate only, since no other evidence of the tumour can be found. 0 = No, 1 = Yes	Number	
SIMD Decile	simdyear_sc_decile	Where year = 2004,2006,2009,2012 and 2016. Scottish Index of Miltiple Deprivation Decile.	Number	
SIMD Quintile	simd2016_sc_quintile	Where year = 2004,2006,2009,2012 and 2016. Scottish Index of Miltiple Deprivation Quintile.	Number	

V1.0 September 2020

Urban Rural (6 score)	ur6_year	Where year = 2003,…2016. Scottish Government 6 fold urban/rural classification. 1 = "Large urban area" 2 = "Other urban area" 3 = "Accessible small town" 4 = "Remote small town" 5 = "Accessible rural" 6 = "Remote rural"	Number	
Urban Rural (8 score)	ur8_year	Where year = 2003,…2016. Scottish Government 8 fold urban/rural classification.	Number	
Tumour information				
ICD10 Code	icd10s_cancer_site	The anatomical site of origin of the primary tumouor, format ICD10	Text	
ICD02 Code	icdo2_icdo2	The anatomical site of origin of the primary tumouor, format ICD02	Text	
ICD03 Code	type_icdo3	The histology of the tumour and comprises the first four digits of the ICD03 morphology	Text	
Grade classification system	grade_classification	The classification system used for grading the tumour. 1 = "Grading for Breast Cancer" 2 = "ICD0/UICC grading system" 3 = "Gleason Score (prostate)" 5 = "Fuhrman Nuclear Grade" 6 = "WHO grade for brain and CNS tumours" 8 = "Other" 9 = "Not determined/not stated/not appicable"	Number	
Differentiation	grade_cell_type	Indicates the degree of differentiation of malignant tumours. 0 - 10, G1, G2, GX.	Text	
Side	side	This indicates the side or laterality in the case of paired organs. 0 = "Not applicable" 1 = "Right" 2 ="Left" 3 = "Bilateral" 9 = "Not known"	Number	
Detection method	method_1st_detection	Indicates how the tumour was first detected. 1 = "Screening examination" 2 = "Incidental finding" 3 = "Clinical presentation" 4 = "Incidental finding at autopsy" 5 = "Interval Cancer" 8 = "Other" 9 = "Not known"	Number	
Most valid base of diagnosis	mvb_diag	Most valid base of diagnosis: indicates the method judged to have provided or validated the diagnosis during the course of the illness. 1 = "Clinical only" 2 = "Clinical investigation (including x-ray, ultrasound, etc." 3 = "Exploratory surgery/endoscopy/autopsy (without concurrent or previous histology)" 4 = "Specific biochemical and/or immunological tests" 5 = "Cytology (including blood film or bone marrow aspirate)" 6 = "Histology of metastasis" 7 = "Histology of primary" 8 = "Autopsy with concurrent or previous histology" 9 = "Not known" 10 = "Death certificate"	Number	
Microscopic confirmation	hist_ver	Microscopic confirmation of the histological or cytological diagnosis. 1 = Verified, 2 = Not Verified.	Number	
Microinvasive	microinvasive	Degree of invasion which is not associated with any risk of nodal metastasis and is sufficiently small to treat by local or conservative means. 0 = No, 1 = Yes, 9 = Not known	Number	
Clinical T Stage	stage_clinical_t	Indicates the extent of the spread of the tumour at diagnosis in terms of clinical findings. Stage is associated with invasive tumours only. T-size/extent of primary tumour based on clinical examination +- imaging	Text	
Clinical N Stage	stage_clinical_n	Condition of regional lymph nodes/glands based on clinical examination +- imaging.	Text	

94

Clinical M Stage	stage_clinical_m	Indicates distant metastates	Text	
Pathologic T Stage	stage_pathologic_t	Indicates the extent of the spread of the tumour at diagnosis in terms of the pathalogical findings. Stage is associated with invasive tumours only. T-size/extent of primary tumour based on clinical examination +- imaging. Breast and lung	Text	
Pathologic N Stage	stage_pathologic_n	Condition of regional lymph nodes/glands based on clinical examination +- imaging	Text	
Pathologic M Stage	stage_pathologic_m	Indicates distant metastates	Text	
Dukes Stage (full)	stage_colorectal	Indicates the extent of spread of the invasive tumour at diagnosis in terms of the pathological and/or clinical findings for Socrates. Extent of primary tumour for Colorectal Cancer Dukes staging is primarily based on histological findings. Stages: A,B,C,C1,C2,D, Unknown.	Text	
Dukes Stage	dukes_stage	Duke's staging of colorectal cancer. 1 = "Duke's Stage A" 2 = "Duke's Stage B" 3 = "Duke's Stage C" 4 = "Duke's Stage D" 9 = "Duke's Stage Unknown"	Number	Derived
CRC ICD10 Codes	crc_icd10	CRC ICD-10 code. 0 = "Other" 1 = "C18:Colon" 2 = "C19: Rectosigmoid junction" 3 = "C20: Rectum" 4 = "C21:Anus/Anal Canal"	Number	
CRC Type	crc_type	Type of colorectal cancer. 1 = "Colon/Sigmoid" 2 = "Rectal"	Number	Derived from crc_icd10
CRC Flag	crc_flag	Diagnosed with colon, sigmoid or rectal cancer. 0 = No, 1 = Yes	Number	Derived from crc_type
Other cancer flag	other_cancer_flag	Patient had a non-crc diagnois during the study period. 0 = No, 1 = Yes	Number	Derived from icd10s_cancer_site
Number of other cancers	count_other_cancer	Number of non-crc diagnoes the patient had during the study period	Number	
Pervious cancer	previous_cancer	Patient had a prior cancer diagnosis (pre-2006)	Number	
Previous CRC cancer	previous_crc_cancer	Patient had a prior CRC cancer diagnosis (pre-2006)	Number	
Treatment				
Nodes examined	nodes_examined	Pathological nodal status -Indicates what regional lymph nodes were examined. 0 = "No regional lymph nodes removed or aspirated" 1 = "Aspiration or biopsy of regional lymph node" 2 = "Sentinel lymph node biopsy" 3 = "Regional lymph node dissection" 4 = "Not known"	Number	
Number of nodes examined	no_of_nodes_examined	Indicates how many of the regional lymph nodes were examined.	Number	
Positive nodes	positive_nodes	Indicates if any of the regional lymph nodes were positive. 1 = Yes, 0 = No, 9 = Not known	Number	

Number of positive nodes	no_positive_nodes	Pathological nodal status -Indicates how many of the regional lymph nodes were positive.0 onwards, Not Known.	Text	
Diagnosis institution	hosp_gp_diag	Institution code in which the diagnosis was first made.	Text	
Health Board of diagnosis	smr06_health_board	Health board in which the diagnosis was first made according to SMR06	Text	Derived from hosp_gp_diag
Referred to radiotherapy	ref_to_rad	Referred to radiotherapy department 1 = Yes, 0 = No, 7 = Planned, 9 = Not known	Number	
Radiotherapy	treated_with_rad	Treated with radiotherapy. 1 = Yes, 0 = No, 7 = Planned, 9 = Not known	Number	
Type of radiotherapy	type_radio_type	Type of radiotherapy administered from 01.01.2006 (variable added to the file during June 2018)	Number	
Radiotherapy to primary site	rad_to_primary	Radiotherapy to primary site. 1 = Yes, 0 = No, 9 = Not known	Number	
Radiotherapy to metastases	rad_to_mets	Radiotherapy to Metastases. 1 = Yes, 0 = No, 9 = Not known	Number	
Other radiotherapy	rad_other	Radiotherapy treatment Other such as organ ablation or prophylaxis. 1 = Yes, 0 = No, 9 = Not known	Number	
Date of first radio therapy	dor_fmt	Date of first radiotherapy treatment: YYYY/MM/DD	Date	
Institution radiotherapy	hosp_1st_rad	Institution code/practice code of hospital/GP practice of first radiotherapy (if applicable)	Text	
Chemotherapy	chemo	Indicates if the patient has had systemic chemotherapy treatment. 1 = Yes, 0 = No, 7 = Planned, 9 = Not known	Number	
Date of first chemo	doc_fmt	Date of first chemo treatment: YYYY/MM/DD	Date	
Institution chemotherapy	hosp_gp_1st_chemo	Institution code/practice code of hospital/GP practice of first chemotherapy (if applicable)	Text	
Hormone therapy	horm_therapy	Indicates if the patient has had hormone therapy treatment. 1 = Yes, 0 = No, 7 = Planned, 9 = Not known	Number	
Date of first hormone therapy	doh_fmt	Date of first hormone treatment: YYYY/MM/DD	Date	
Institution hormone therapy	hosp_gp_1st_hormone_therapy	Institution code/practice code of hospital/GP practice that initiated the first hormone therapy (if applicable)	Text	
Surgery	surgery	Indicates if the patient has been treated with surgery. 1 = Yes, 0 = No, 7 = Planned, 9 = Not known	Number	
Date of first surgery	dos_fmt	Date of first surgical treatment: YYYY/MM/DD	Date	
Institution surgery	hosp_gp_1st_surgery	Institution code/practice code of hospital/GP practice where a surgical treatment was first carried out (if applicable)	Text	

Palliative surgery	palliative_surgery	Treated with palliative surgery (variable added to the file during June 2018). 1 = Yes, 0 = No, 7 = Planned, 9 = Not known	Number	
Date of first palliative surgery	dops_fmt	Date of first palliative surgical treatment: YYYY/MM/DD	Date	
Institution palliative surgery	hosp_gp_1st_pal_su rgery	Institution code/practice code of hospital/GP practice where first palliative surgery was carried out(if applicable)	Text	
Immunotherapy	type_immun_type	Indicates if immunitherapy/biotherapy was administered. From 01.01.1997 (variable added to the file during June 2018) 1 = Yes, 0 = No, 7 = Planned, 9 = Not known	Number	
Date of first immunotherapy	doit_fmt	Date of first immunotherapy type treatment: YYYY/MM/DD	Date	
Institution immunotherapy	hosp_gp_1st_immun _type	Institution code/practice code of hospital/GP practice that initiated the first immunotherapy (if applicable)	Text	
Other therapy	other_therapy	Patient treated with other therapy. 1 = Yes, 0 = No, 7 = Planned, 9 = Not known	Number	
Date of first other therapy	doot_fmt	Date of first other therapy treatment: YYYY/MM/DD	Date	
Institution other therapy	hosp_gp_1st_other_t herapy	Institution code/practice code of hospital/GP practice that initiated the first other therapy (if applicable)		
Other therapy	type_other_therapy	Indicates the type of other therapy treatment carried out.	Text	
Objectives of treatment	therapy_objectives	Objective of treatment. 1 = "Curative intent" 2 = "Non-curative intent (palliative)" 9 = "Not Known"	Number	

NRS deaths data set (2006-2019)				
Data Item	**Variable Name**	**Description of field content**	**Format**	**Further info**
Identifiers				
Master ID	master_index	Pseudononymised person ID	Text	
Patient information				
Date of birth	dob_fmt	Date of Birth: MMYYYY	Date	
Date of death	dod_fmt	Date of death: DDMMYYYY	Date	
Gender	sex	Sex 1 = Male; 2 = Female	Number	
Age died	age_died	Age at death.	Number	Derived from dob_fmt and dod_fmt
SIMD Decile	simdyear_sc_decile	Where year = 2004,2006,2009,2012 and 2016. Scottish Index of Miltiple Deprivation Decile.	Number	

V1.0 September 2020

SIMD Quintile	simd2016_sc_quintile	Where year = 2004,2006,2009,2012 and 2016. Scottish Index of Miltiple Deprivation Quintile.	Number	
Urban Rural (6 score)	ur6_year	Where year = 2003,…2016. Scottish Government 6 fold urban/rural classification. 1 = "Large urban area" 2 = "Other urban area" 3 = "Accessible small town" 4 = "Remote small town" 5 = "Accessible rural" 6 = "Remote rural"	Number	
Urban Rural (8 score)	ur8_year	Where year = 2003,…2016. Scottish Government 8 fold urban/rural classification.	Number	
Details of death				
Underlying cause of death	underlying_cause_of _death	The disease or injury which initiated the chain of morbid events leading directly to death, or the accident/act which produced the fatal injury. ICD10 codes.	Text	
Other causes of death	cause_of_death_code_i	Other causes of death mentioned on the death certificate, after the primary cause of death. ICD10 codes. Where i = 0,…,9.	Text	
Duration of illness in years	duration_of_illness_years_1i	Approximate interval between onset of illness and death in years. Where i= a,b,c,d.	Number	
Duration of illness in months	duration_of_illness_months_1i	Approximate interval between onset of illness and death in months. Where i= a,b,c,d.	Number	
Duration of illness in days	duration_of_illness_days_1i	Approximate interval between onset of illness and death in days. Where i= a,b,c,d.	Number	
Place of death	place_of_death	Place where death occurred. 0 = "Institution invalid or irrelevant" 1 = "NHS Hospital" 2 = "Home/Private Address" 3 = "Hospice" 4 "Private care homes and care homes" 5 "Homes for the elderly" 6 "Private hospital" 7 "Other"	Number	
Institution of death	institution	The institution code for where the death occurred.	Text	
Primary household occupation	occupation	Primary household occupation code.	Text	
Primary household occupation group	Major_occ_group	Primary household major occupation group. 0 = "Large employers and higher managerial occupations" 1 = "Higher professional occupations" 2 = "Lower managerial and professional occupations" 3 = "Intermediate occupations" 4 = "Small employers and own account workers" 5 = "Lower supervisory and technical occupations" 6 = "Semi-routine occupations" 7 = "Routine occupations" 8 = "Never worked and long term unemployed" 9 = "Students, not stated or not classifiable"	Number	Derived from occupation
Occupation	deceased_occupation_code	Classification of the deceased persons occupation code.	Text	
Major occupation group	dec_major_occ_group	Deceased persons major occupation group. 0 = "Large employers and higher managerial occupations" 1 = "Higher professional occupations" 2 = "Lower managerial and professional occupations" 3 = "Intermediate occupations" 4 = "Small employers and own account workers" 5 = "Lower supervisory and	Number	Derived from deceased_occupation_code

Data Item	Variable Name	Description of field content	Format	Further info
		technical occupations" 6 = "Semi-routine occupations" 7 = "Routine occupations" 8 = "Never worked and long term unemployed" 9 = "Students, not stated or not classifiable"		
Employment status	deceased_employm ent_status	The employment status of the deceased. 0 = "Other- student, unemployed, not available, etc" 1 = "Employee, apprentice, armed forces- other rank etc" 2 = "Manager, superintendent, armed forces- officer etc" 3 = "Supervisor, foreman, charge hand etc" 4 "Self-employed- with employees" 5 "Self-employed- without employees"	Number	
Colorectal cancer related death	crc_death	An indicator that shows if any of the causes of death mentioned on the death certificate were due to colorectal cancer. 0 = Non-CRC death, 1 = CRC death	Number	
Country of residence	country_of_residenc e	Country of residence code for the deceased person. IS03166 codes.	Text	
Scottish resident	scottish_resident	Indicates if the deceased was a Scottish resident. 0 = No, 1 = Yes	Number	Derived from country_of_reside nce
Health Board	deaths_health_board	Health board where the death occurred. A= "Ayrshire and Arran" B = "Borders" C = "Argyll and Clyde" F = "Fife" G = "Greater Glasgow" H = "Highland" L ="Lanarkshire" N = "Grampian" R = "Orkney" S = "Lothian" T = "Tayside" V = "Forth Valley" W = "Western Isles" Y = "Dumfries and Galloway" Z = "Shetland"	Text	Derived from institution.

SMR01 data set (1997-2018)				
Data Item	**Variable Name**	**Description of field content**	**Format**	**Further info**
Identifiers				
Master ID	master_index	Pseudononymised person ID	Text	
Patient information				
Date of birth	dob_fmt	Date of Birth: MMYYYY	Date	
Gender	sex	Sex 0 = Not Known; 1 = Male; 2 = Female; 9 = (includes not stated by patient, or not recorded)	Number	
SIMD Decile	simdyear_sc_decile	Where year = 2004,2006,2009,2012 and 2016. Scottish Index of Miltiple Deprivation Decile.	Number	
SIMD Quintile	simd2016_sc_quintil e	Where year = 2004,2006,2009,2012 and 2016. Scottish Index of Miltiple Deprivation Quintile.	Number	
Urban Rural (6 score)	ur6_year	Where year = 2003,…2016. Scottish Government 6 fold urban/rural classification. 1 = "Large urban area" 2 = "Other urban area" 3 = "Accessible small town" 4 = "Remote small town" 5 = "Accessible rural" 6 = "Remote rural"	Number	

V1.0 September 2020

Urban Rural (8 score)	ur8_year	Where year = 2003,…2016. Scottish Government 8 fold urban/rural classification.	Number	
Episode level information				
Date of admission	doa_fmt	Date of admission: YYYYMMDD	Date	
Date of discharge	dodis_fmt	Date of discharge: YYYYMMDD	Date	
Year of admission	yoa	Year of admission	Number	Derived from doa_fmt
Health Board	smr01_health_board	Health Board. A= "Ayrshire and Arran" B = "Borders" C = "Argyll and Clyde" F = "Fife" G = "Greater Glasgow" H = "Highland" L ="Lanarkshire" N = "Grampian" R = "Orkney" S = "Lothian" T = "Tayside" V = "Forth Valley" W = "Western Isles" Y = "Dumfries and Galloway" Z = "Shetland"	Text	
Institution type	smr01_institution_type	Institution type. C = "Clinic Premises" H = "NHS Hospital" J = "Joint user hospital" K = "Contractural hospital" V = "Private nursing home, private hospital"	Text	
Type of admission	admission_type	Admission type. 10 ="Routine Admission - no additional detail added" 11= "Routine elective (i.e. from waiting list as planned, excludes planned transfers)" 12 ="Patient admitted on day of decision to admit, or following day, not for medical reasons, but because suitable resources are available" 18 ="Planned transfers" 19 ="Routine Admission - type not known" 20 ="Urgent Admission - no additional detail added" 21 ="Patient delay (for domestic, legal or other practical reasons)" 22= "Hospital delay (for administrative or clinical reasons e.g. arranging appropriate facilities, for tests to be carried out, specialist equipment, etc.)" 30 ="Emergency Admission - no additional detail added" 31= "Patient Injury - Self Inflicted (Injury or Poisoning)" 32 ="Patient Injury - Road Traffic Accident (RTA)" 33 ="Patient Injury - Home Incident (incl. assault or accidental poisoning)" 34 ="Patient Injury - Incident at Work (incl. assault or accidental poisoning)" 35= "Patient Injury - Other Injury (including assault or accidenta poisoning other than in the home or at work)" 36 ="Patient Non-Injury (e.g. stroke, MI, ruptured appendix)" 38 ="Other Emergency Admission (including emergency transfers)" 39= "Emergency Admission - type not known"	Number	
Type of admission group	admission_type_group	Type of admission- higher level grouping. 1 = "Routine admission" 2 ="Urgent admission" 3 ="Emergency admission" 9 = "Admission type unknown"	Number	

Patient category	patient_category	Patient category. 1 = "Amenity" 2= "Paying" 3 = "NHS" 4 = "Overseas visitor - liable to pay for treatment" 5 = "Overseas visitor - not liable to pay" 8 = "Other (including Hospice)"	Number	
Continuous inpatient stay	cis_marker	Continuous inpatient stay marker- CIS is the unbroken period of care that a patient spends as an inpatient	Number	
Specialty	specialty	The division of medicine or dentistry covering a specific area of clinical activity	Text	
Significant facility	significant_facility	Type of clinical facility which is identified for clinical and/or costing purposes	Text	
Discharge/transfer to	discharge_transfer_to	Type of location a patient is discharged/transferred to following an episode of care	Text	
Discharge/transfer to group	dis_trans_to_group	Main grouping for discahrge/transfer to group. 1 = "Patient died" 10 = "Private residence" 20 = "Institution" 30 = "Temporary place of residence" 40 = "Transferred within the same health board/health care provider" 50 = "Transferred to another health board/health care provider" 60 = "Other type of location"	Number	Derived from discharge_transfer_to
Discharge type group	dis_type_group	Main grouping for discharge type. 10 = "Regular discharge" 20 = "Irregular discharge" 40 = "Death"	Number	Derived from discharge_type
Discharge type	discharge_type	Indicates whether a discharge from an inpatient or day case epsiode is regular,irregular or due to patient death. 10= "Regular Discharge - no additional detail added" 11 = "Discharge from NHS inpatient/day case care" 12 ="Transfer within the same Health Board/ Health Care Provider" 13 = "Transfer to another Health Board/ Health Care Provider" 18 = "Other type of regular discharge" 19= "Regular discharge - type not known" 20 = "Irregular Discharge - no additional detail added" 21= " Patient discharged himself/herself against medical advice" 22= "Patient discharged by relative" 28 Other type of irregular discharge" 29= Irregular discharge - type not known" 40 ="Death - no additional detail added" 41 =Death - Post Mortem"	Number	
HRG	hrg	Healthcare resource group.	Text	
Length of stay	length_of_stay	Length of stay in hospital days (at episode level)	Number	
Medical conditions and treatment				
Main medical condition	main_condition	Main medical (or social) condition managed/investigated during the patient's stay	Text	

Other medical condition	other_conditition_i	Where i= 1,…,5. The i'th other medical condition managed/investigated during the patient's stay after the main condition	Text	
Main condition colorectal cancer	main_condition_crc	Indicates the type of CRC if the main condition was CRC. 0 = "Non-CRC" 1 = "C18:Colon" 2 = "C19: Rectosigmoid junction" 3 = "C20: Rectum"	Number	
Any condition colorectal cancer	any_condition_crc	Indicates the type of CRC if the any condition was CRC. 0 = "Non-CRC" 1 = "C18:Colon" 2 = "C19: Rectosigmoid junction" 3 = "C20: Rectum"	Number	
Main operation	main_operation	Main operation carried out during patient stay (OPCS4)	Text	
Other operations	other_operation_i	Where i= 1,…,3. Other operation carried out during patient stay (i'th after main)	Text	
OPCS4 Code	opcode2	First OPCS4 code of main_operation	Text	Derived from main_operation
OPCS4 Code 2B	opcode2b	Second OPCS4 code of main_operation	Text	Derived from main_operation
OPCS4 Code Group	opcode_derived	Grouping variable for colorectal operation. 1 = "Major resection" 2 = "Minor resection" 3 = "Bypass/stoma" 4 = "Stent" 5 = "Appendix" 6 = "Liver" 7 = "Other/currently unknown"	float	
OPCS4 Code Group 2B	opcode2b_derived	Grouping variable for colorectal operation where additional info is given for the same operation. 1 = "Major resection" 2 = "Minor resection" 3 = "Bypass/stoma" 4 = "Stent" 5 = "Appendix" 6 = "Liver" 7 = "Other/currently unknown"	float	
Acute myocardial infarction	ami	Acute myocardial infarction. 0 = No, 1 = Yes	Number	
Congestive heart failure	chf	Congestive heart failure. 0 = No, 1 = Yes	Number	
Peripheral vascular disease	pvd	Peripheral vascular disease. 0 = No, 1 = Yes	Number	

V1.0 September 2020

Cerebrovascular disease	cevd	Cerebrovascular disease. 0 = No, 1 = Yes	Number	
Dementia	dementia	Dementia. 0 = No, 1 = Yes	Number	
COPD	copd	Chronic obstructive pulmonary disease. 0 = No, 1 = Yes	Number	
Rheumatoid disease	rheumd	Rheumatoid disease - Connective tissue disease. 0 = No, 1 = Yes	Number	
Peptic ulcer disease	pud	Peptic ulcer disease. 0 = No, 1 = Yes	Number	
Mild liver disease	mld	Mild liver disease. 0 = No, 1 = Yes	Number	
Diabetes no complications	diab	Diabetes no complications. 0 = No, 1 = Yes	Number	
Diabetes w complications	diabwc	Diabetes w complications. 0 = No, 1 = Yes	Number	
Hemiplegia or paraplegia	hp	Hemiplegia or paraplegia. 0 = No, 1 = Yes	Number	
Renal disease	renal	Renal disease. 0 = No, 1 = Yes	Number	
Cancer (any malignancy)	cancer	Cancer (any malignancy). 0 = No, 1 = Yes	Number	
Moderate or severe liver disease	msld	Moderate or severe liver disease. 0 = No, 1 = Yes	Number	
Metastatic solid tumour	metacancer	Metastatic solid tumour. 0 = No, 1 = Yes	Number	
HIV	hiv	HIV. 0 = No, 1 = Yes	Number	

QPI data set (2013-2018)				
Data Item	**Variable Name**	**Description of field content**	**Format**	**Further info**
Identifiers				
Master ID	master_index	Pseudononymised person ID	Text	

Patient information				
Date of birth	dob_fmt	Date of Birth: MMYYYY	Date	
Sex	sex	Sex. 1 = Male; 2 = Female;	Number	
Date of death	dod_fmt	Date of Death: DDMMYYYY	Date	
Diagnosis				
Source of cancer referral	mrefer	Source of Cancer Referral. 1 "Primary care" 2 "Screening" 3 "Incidental" 4 "Review clinic" 5 "Cancer genetic clinic" 6 "Self-referral A&E" 7 "GP direct to hospital" 8 "Previous GP but subsequent to hospital" 11 "Primary care clinician (dental)" 12 "Private healthcare" 13 "Other" 99 "Not recorded"	Number	
Location of diagnosis	hosp	Location of diagnosis	Text	
Health board of diagnosis(QPI)	qpi_health_board	Health Board according to QPI hosp of diagnosis. A= "Ayrshire and Arran" B = "Borders" C = "Argyll and Clyde" F = "Fife" G = "Greater Glasgow" H = "Highland" L ="Lanarkshire" N = "Grampian" R = "Orkney" S = "Lothian" T = "Tayside" V = "Forth Valley" W = "Western Isles" Y = "Dumfries and Galloway" Z = "Shetland"	Derived from hosp	
Date of diagnosis	diagdate_fmt	Date of Diagnosis : DDMMYYYY	float	
Date of histological diagnosis	hdiag	Date of Histological Diagnosis	long	
Most valid basis of diagnosis	valid	Most valid basis of diagnosis. 1 ="Clinical only" 2 ="Clinical Investigation" 3 ="Exploratory surgery/endoscopy/autopsy" 4= "Tumour specific markers" 5= "Cytology" 6= "Histology of metastasis" 7= "Histology of primary" 99= "Not known"	Number	
Staging/Imaging				
Staging investigations complete	sinvest	Staging investigations complete. 1 "Complete CTCAP" 2 "Complete CTCAP and MRI" 3 "Incomplete" 4 "Incomplete - Contraindications" 95 "Patient refused" 96 "Not applicable" 99 "Not recorded"	byte	
Data staging investigations complete	sinvestdate_fmt	Date Staging Inv completed: DDMMYYYY	Date	
Large bowel imaging	lbtype	Large bowel imaging. 1 "Yes Colonoscopy or CT Colonography" 2 "Incomplete" 3 "Not performed" 4 "Incomplete due to obstructing tumour" 94 "Patient died before treatment" 95 "Patient refused" 96 "Not applicable" 99 "Not recorded"	Number	
Date of large bowel imaging	lbdate_fmt	Date of Large Bowel Imaging : DDMMYYYY	Date	

V1.0 September 2020

Treatment				
Date of discussion by MDT	mdtdate_fmt	Date discussed by multidisciplinary team (MDT): DDMMYYYY	Date	
Seen by stoma nurse	stomanurse	Seen by Stoma Nurse. 1= "Yes" 2 ="No" 95= "Patient refused" 96 ="Not applicable" 99 ="Not recorded"	Number	
Date seen by stoma nurse	stomandate_fmt	Date of Stoma nurse : DDMMYYYY	Date	
Stoma site marked	stomamark	Stoma Site Marked Pre-op. 1= "Yes" 2 ="No" 95= "Patient refused" 96 ="Not applicable" 99 ="Not recorded"	Number	
Date of first treatment	firsttreatdate_fmt	Date of first treatment: DDMMYYYY	Date	
Type of first cancer treatment	firsttreatmode	Type of First Cancer Treatment. 1= "Surgery" 2 ="Radiotherapy" 3 ="Chemotherapy" 4= "Chemoradiotherapy" 5= "Endoscopic" 7= "Supportive Care Only" 11= "Other therapy" 94= "Patient died before treatment" 95 ="Patient refused treatment" 99= "Not recorded"	Number	
Date of definitive treatment	deftreatdate_fmt	Date of definitive treatment : DDMMYYYY	Date	
ASA status	asa	American Society of Anaesthesiologists (ASA) status. The ASA PS classification globally assesses the degree of "sickness" or "physical state" prior to selecting the anaesthetic or prior to performing surgery 1= "Normal healthy patient" 2 ="Mild systemic disease" 3 ="Severe systemic disease" 4 ="Severe systemic disease constant threat to life" 5 ="Moribund" 6 ="Brain-dead" 96 ="Not applicable" 99 ="Not recorded"	Number	
Location of surgery	hospsurg	Location of Surgery	Text	
OPCS4 Code	opcode2	Final Definitive Surgery Performed CRC	Text	
OPCS4 Code 2B	opcode2b	Final Definitive Surgery Performed CRC 2b	Text	
OPCS4 Code Group	opcode_derived	Grouping variable for colorectal operation. 1 = "Major resection" 2 = "Minor resection" 3 = "Bypass/stoma" 4 = "Stent" 5 = "Appendix" 6 = "Liver" 7 = "Other/currently unknown"	Number	
OPCS4 Code Group 2B	opcode2b_derived	Grouping variable for colorectal operation where additional information is given for the same operation. 1 = "Major resection" 2 = "Minor resection" 3 = "Bypass/stoma" 4 = "Stent" 5 = "Appendix" 6 = "Liver" 7 = "Other/currently unknown"	Number	
Total Mesorectal Excision	texcision	Total Mesorectal Excision. 1 ="Yes" 2 ="No" 96 ="Not applicable" 99= "Not recorded"	Number	

Type of surgical approach	surgappr	Surgical Approach. 1 ="Open" 2 ="Laparoscopic completed" 3 ="Laparoscopic converted" 4 ="TEM" 5 "TART" 96 ="Not applicable" 99= "Not known"	Number	
Date of definitive surgery	finsurgdate	Date of Definitive Surgery CRC: DDMMYYYY	Date	
Presentation type	present	Presentation type. 1 ="Elective routine" 2= "Emergency" 96= "Not applicable" 99= "Not recorded"	Number	
Anastomotic leak	anasleak	Anastomotic Leak. 1= "Yes" 2 ="No anastomotic leak" 96 ="Not applicable" 99 ="Not known"	Number	
Intent of surgery	opintent	Intent of Surgery. 1= "Curative" 2= "Palliative" 96 ="Not applicable" 99 ="Not recorded"	Number	
Re-operation	reoper	Re-operation. 1 ="Yes" 2= "No" 96= "Not applicable" 99 ="Not recorded"	Number	
Extramural venous invasion	extra	Extramural venous invasion. 1 ="Present" 2 ="Not present" 96 ="Not applicable" 99= "Not recorded"	Number	
Circumferential margin involved	circmargin	Circumferential margin involved. 1 ="Involved" 2 ="Not involved" 96= "Not applicable" 99 ="Not recorded"	Number	
Location of SACT	hospsact	Location of SACT	Text	
Location of radiotherapy	hospradio	Location of radiotherapy	Text	
Neo-Adjuvant Oncology Treatment Type.	neoonc	Neo-Adjuvant Oncology Treatment Type. 1 ="Short Course RT" 2= "Long Course RT with chemo" 3 ="Long Course RT alone" 4 ="Chemotherapy" 80 ="Patient died before radiotherapy" 81 ="Patient died before SACT" 82 ="Patient died before chemoRT" 83 ="Patient refused RT" 84 "Patient refused SACT" 85 ="Patient refused chemoRT" 86 "RT contraindicated" 87 ="Chemotherapy contraindicate" 88 ="ChemoRT contraindicated" 96 ="Not applicable" 99 ="Not recorded"	Number	
Date Neo-adjuvant oncology treatment started	neoadjdate_fmt	Date Neo-adjuvant oncology treatment started : DDMMYYYY	Date	
Date Neo-adjuvant oncology treatment completed	neoadjcom_fmt	Date Neo-adjuvant oncology treatment completed : DDMMYYYY	Date	
Primary/Palliative/ Adjuvant Oncology Treatment Type	adjonc	Primary/Palliative/Adjuvant Oncology Treatment Type. 1 ="Adjuvant Long Course RT with chemotherapy" 2= "Adjuvant chemotherapy" 3= "Adjuvant RT" 4= "Primary Chemotherapy" 5= "Primary/Radical RT" 6 ="Palliative RT" 7 ="Palliative Chemotherapy" 8= "Biological Therapy" 9 ="Palliative Chemoradiotherapy" 80 ="Patient died before radiotherapy" 81 ="Patient died before SACT" 82 ="Patient died before chemoRT" 83 ="Patient refused RT"	Number	

		84 ="Patient refused SACT" 85 ="Patient refused chemoRT" 96 ="Not applicable" 99= "Not recorded"		
Date Primary/Palliative/ Adjuvant Oncology Treatment started	adjoncdate_fmt	Date Primary/Palliative/Adjuvant Oncology Treatment started : DDMMYYYY	Date	
Date Primary/Palliative/ Adjuvant Oncology Treatment completed	adjcom_fmt	Date Primary/Palliative/Adjuvant Oncology Treatment completed : DDMMYYYY	Date	
Health Board	hbyear	Health Board code using various year codes. Where year = 2006, 2014, 2018, 2019.	Text	
Tumour characteristics				
Site of origin of primary tumour	site	Site of Origin of Primary Tumour. 1 ="Caecum" 2 ="Currently unknown await update" 3= "Ascending colon" 4= "Hepatic flexure" 5 ="Transverse colon" 6 ="Splenic flexure" 7= "Descending colon" 8 ="Sigmoid colon" 9 ="Overlapping lesion of colon" 10 ="Colon, unspecified" 11= "Rectum" 12 ="Not recorded"	Number	
Circumferential margin involved	circmargin	Circumferential margin involved. 1 ="Involved" 2 ="Not involved" 96= "Not applicable" 99 ="Not recorded"	Number	
Grade of differentiation	different	Grade of differentiation. 1= "Well/moderate" 2 ="Poor" 3= "Not assessable" 96 ="Not applicable" 99= "Not known"	Number	
Number of lymph nodes examined	lnexamine	Final total number of lymph nodes examined microscopically	Number	
Number of lymph nodes involved	lninvolve	Number of lymph nodes involved	Number	
TNM Tumour	finalt	TNM tumour	Number	
TNM Nodes	finaln	TNM nodes	Number	
TNM Metastasis	finalm	TNM Metastasis	Number	
Dukes staging	dukes	Dukes stage. 1 ="Dukes A" 2 ="Dukes B" 3= "Dukes C1" 4 ="Dukes C2" 5= "Dukes D" 96= "Not applicable" 99 ="Not recorded"	Number	
Dukes staging derived from TNM	dukes_derived_qpi	Dukes stage derived from TNM staging. 1 ="Dukes A" 2 ="Dukes B" =3 "Dukes C" 4 ="Dukes D"	Number	Derived from finalt, finaln, finalm
Cancer Network specific				
Cancer Network	location	Indicates regional cancer network. 1 = SCAN, 2 = WoScan, 3 = NoSCAN	Number	Derived

SCAN	SCAN_flag	Indicates if the record is from SCAN. 0 = No, 1 = Yes	Number	
WoSCAN	Wos_flag	Indicates if the record is from Wo-SCAN. 0 = No, 1 = Yes	Number	
Distance from anal verge (SCAN only)	analverge_SCAN	Distance from Anal Verge	Number	
CT Chest Result (SCAN only)	cxr_SCAN	CT Chest Result. 1 ="No metastases" 2 ="Metastases" 3 ="Equivocal" 4 ="Not performed" 95 ="Patient refuses investigation" 96 ="Not applicable" 99 ="Not recorded"	Number	
Date of CT Chest Result (SCAN only)	xdate_SCAN	Date of CT Chest Result	Date	
Liver imaging Result (SCAN only)	liver_SCAN	Liver imaging Result. 1 ="No metastases" 2 ="Metastases" 3 ="Equivocal" 4= "Not performed" 95 ="Patient refused investigation" 96 ="Not applicable" 99 ="Not recorded"	Number	
Date of liver imaging completed (SCAN only)	liverdate_SCAN_fmt	Date of liver imaging SCAN completed : DDMMYYYY	Date	
Circumferential Resection Margin Predicted (SCAN only)	cmarginpredict_SCAN	Circumferential Resection Margin Predicted. 1 ="Clear" 2 ="Threatened" 3 ="Involved" 96 ="Not applicable" 98= "Not assessed" 99= "Not recorded"	Number	

Chemocare data set (2013-2018)				
Data Item	**Variable Name**	**Description of field content**	**Format**	**Further info**
Identifiers				
Master ID	master_index	Pseudononymised person ID	Text	
Patient information				
Gender	sex	Sex of patient. 1 = Male; 2 = Female	Number	
Patient height	height	Patient height in meters	Float	
Patient weight	weight	Patient weight in kg	Float	
Body surface area	body_surface_area	Calculation of body surface area based on height and weight (generated by Chemocare)	Float	
Intention of treatment	intention	Intention of treatment (as in Chemocare)	String	

Intention of treatment grouping	intention_group	Intention of treatment. 1 ="Adjuvant" 2 ="Curative" 3 ="Radical" 4 ="Neo-Adjuvant" 5 ="Palliative" 6 ="Peri-op" 7= "Durable Remission" 8 ="Day Case" 9 ="MAI" 10 ="INT" 11= "CON" 99= "Unknown"	Number	Derived from intention
Appointment date	appointment_date	Appointment date chemotherapy drug was administered: DDMMYYYY	Date	
Cycle number	cycle_number	Cycle of chemotherapy (generated by Chemocare)	Number	
Day number	day_number	Day of the cycle of chemotherapy (generated by Chemocare)	Number	
Drug name	drug_name	Drug name for any drug administered during chemotherapy appointment.	Text	
Drug dose	drug_dose	Standard drug dose per meter squared for specific drug or dose band.	Number	
Required dose	required_dose	Actual dose delivered to patient.	Number	
Diagnosis	diagnosis	Diagnosis (original text from Chemocare)	Text	
Diagnosis grouping	Diagnosis_group	Diagnosis to differentiate colon and rectum, and may indicate the intention of treatment e.g. palliative, curative and line of treatment in metastatic setting e.g. 1st, 2nd, 3rd. . 1 ="Colon" 2 ="Rectal" 3= "Adjuvant Colon" 4 ="Adjuvant Rectum" 5 ="Adj rectum after downstaging" 6 ="Adj Rectal after primary" 7= "Adj rectum post resection" 8 "Pal Colon 1st" 9 = "Pal Colon 2nd" 10= "Pall Colon 3rd" 11= "Pall Colon 4th" 12= "Pall Rectum 1st" 13= "Pall Rectum 2nd" 14 ="Pall Rectum 3rd" 15 ="Pall Rectum 4th" 16= "Neoadj rectum" 17 ="Neoadj colon liver" 18 "Neoadj rectum liver" 19= "Periop colon" 20= "Periop rectum" 21= "Pseudomyxoma 1st" 22 ="Pseudomyxoma 2nd" 23= "Anal" 24 ="Misc" 25= "LowerGI"	Number	Derived from diagnosis
Regime	regime	Description of drug regimen given to patient.	Text	
Cancer Network specific				
Cancer Network	location	Location. 1 ="SCAN" 2 ="WoSCAN" 3= "Grampian" 4 ="Highland" 5 ="Tayside"	Number	
Protocol (SCAN only)	protocol_SCAN	Further explanation about drug regimen. Available for SCAN only.	Text	
Drug type (SCAN only)	drug_type_SCAN	Type of drug e.g. chemotherapy, anti-sickness, anti-diahoerea etc.	Text	
Hospital (SCAN only)	hospital_SCAN	Hospital. BGH = Borders General Hospital, DRI = Dumfries and Galloway Royal Infirmary, SJH = St John's Hospital, VHF = Victoria Hospital, WGH = Western General Hospital	Text	
Method of dose calculation (WoSCAN. NoSCAN only)	calculation	Method of dose calculation. 1 ="Flat" 2 ="Glomerular Filtration Rate" 3 ="Surface Area" 4 ="Weight" 99= "Unknown"	Number	
Unit of drug (NoSCAN only)	unit	Unit of drug. 1 ="Application" 2= "Capsule" 3 ="Tablet" 4 ="Sachet" 5 ="Drop" 6 ="Gram" 7 ="Mg" 8 ="Mcg" 9 ="IU" 10 ="MU" 11= "Ml" 12= "Mmol" 13 ="Unit" 99= "Unknown"	Number	

109

Date BSA recorded (WoSCAN and Grampian only)	date_bsa_recorded	Date body surface area calculated and updated.	Date	
Capped (WoSCAN only)	capped	Capped dose for body surface area. 0 "No" 1 "Yes" 99 "Unknown"	Number	
Duration (WoSCAN only)	duration	Duration of drug prescription e.g. in days or weeks.	Text	
Frequency (WoSCAN only)	freq	Frequency of drug administration e.g. per-day.	Text	
Performance status (WoSCAN, NoSCAN only)	performance_status	Performance status of patient in terms of their general wellbeing and activities of daily living. 0 ="Zero" 1 ="One" 2= "Two" 3 ="Three" 4 ="Four"	Number	
Drug route (WoSCAN and NoSCAN only)	drug_route	The route the drug was administered (e.g oral, IV)	Text	
Drug route grouping (WoSCAN and NoSCAN only)	Drug_route_group	1 = "Intravenous (IV)" 2 = "Oral" 3 = "Intramusculor" 4 = "Sub Cutaneous" 5 = "Topical-skin" 6 = "Eye- topical, drops, intravetrial" 7 = "Mouth wash" 8 = "Buccal" 9 = "Per rectal" 10 = "Other" 99 = "Missing"	Number	Derived from drug_route