CESR Technical Report 1:

The quality and usefulness of the NSW Clinical Cancer Registry Minimum Dataset and Colorectal Dataset Extension for colorectal cancer services research

Cancer Epidemiology and Services Research (CESR)

Sydney School of Public Health

Professor Jane Young

Ms Mikaela Jorgensen

Dr Timothy Dobbins

Professor Michael Solomon

Dec 2012

Acknowledgements

This research was funded by a Cancer Epidemiology Linkage Grant from the Cancer Institute NSW.

Many thanks to the data managers and custodians of each Area Health Service Clinical Cancer Registry, the NSW Central Cancer Registry, NSW Admitted Patients Data Collection, NSW Emergency Department Data Collection and NSW Registry of Births, Deaths and Marriages for releasing the data used in this report and for their valuable comments and suggestions. Thank you to the staff at the Centre for Health Record Linkage (CHeReL) for preparing the data for linkage and for their readiness to assist with queries.

Suggested citation

Young J, Jorgensen M, Dobbins T, Solomon M. *CESR Technical Report 1: The quality and usefulness of the NSW Clinical Cancer Registry Minimum Dataset and Colorectal Dataset Extension for colorectal cancer services research.* Sydney: Sydney School of Public Health, The University of Sydney, Dec 2012.

Table of Contents

List of Tables

List of Figures

ABSTRACT

Colorectal cancer is one of the most common cancers worldwide. Population-based studies of care and outcomes are essential to monitor the uptake of evidence-based treatment guidelines and identify groups most at risk of receiving sub-optimal care or experiencing poor outcomes. With the development of locally-managed Clinical Cancer Registries (ClinCR) in public facilities in many Area Health Services (AHS) since 2006, 'patterns of care' studies which previously relied on the collection of clinical information through time- and resource-intensive surveys or medical record audits now have the potential to be conducted using routinely collected data. However there is little experience with the use of ClinCR data for research. The purpose of this report is to assess the quality, coverage and completeness of ClinCR data for use in colorectal cancer services research, and to assess the feasibility of developing surgical process and outcomes indicators that rely on ClinCR data items.

The records of people with a newly diagnosed cancer of the colon, rectum or anus registered by the NSW Central Cancer Registry (CCR) since 2000 were linked to the following data sources by the Centre for Health Record Linkage (CHeReL): all six AHS ClinCR datasets including the NSW Oncology Group colorectal dataset extension (only collected in two AHS), NSW Admitted Patients Data Collection (APDC), NSW Emergency Department Data Collection (EDDC), and NSW Registry of Births, Deaths and Marriages (RBDM). Data from the AHS ClinCR datasets were harmonised and combined (12,968 records for 5,452 persons), then merged onto the CCR cohort along with the APDC and RBDM. After data cleaning, the final cohort included 37,593 colorectal cancer cases diagnosed between July 2000 and December 2008.

The proportion of the cohort with a ClinCR record (coverage) increased from 73% in 2007 to 85% in 2008 for those having a procedure of interest in a public hospital. This proportion is over 94% for all AHS with a ClinCR in 2008. Demographic, diagnostic and treatment items in the ClinCR have high rates of completeness (mainly over 90%). Staging items have lower levels of completeness (77-84%), and quality of care items are largely only available for less than 60% of people. While discordance of ClinCR items duplicated in other datasets ranges from 4-22% for staging items, diagnostic items and admission-related dates, most demographic items have high concordance and the internal consistency of ClinCR date items is also high for the majority of records (82% with no illogical dates). ClinCR or dataset extension items are required for the calculation of 12 of 21 surgical outcome and evidence-based care indicators that could feasibly be developed from the linked cohort dataset.

The ClinCR and colorectal dataset extension provide a range of data on cancer staging, clinical treatment and care processes that add valuable information to other routinely collected data. While there are limitations in the coverage and quality of the ClinCR and extension data, including difficulty determining whether records have been created for all treatment events, without these datasets the development of population-based surgical indicators from routinely collected sources would be greatly limited. Recommendations resulting from this report include concentrating resources on items that are unique to the ClinCR, increasing coverage to private facilities, determining accuracy of the recording of treatment events, and increasing the collection of the ClinCR and dataset extension to all AHS where they are not currently in operation. Further assessment of the quality of ClinCR data for other tumour types for research purposes is also recommended.

EXECUTIVE SUMMARY

BACKGROUND

Colorectal cancer (CRC) is the second most common registered cancer in New South Wales and is one of the most common cancers worldwide. High-quality evidence exists for treatment approaches in CRC that improves patient outcomes, but these are not always implemented in routine clinical care to all patients.

Population-based studies of treatment, care pathways and outcomes are essential to monitor the uptake of evidence-based guidelines and identify groups most at risk of receiving sub-optimal care or experiencing poor outcomes. However, the collection of clinical information through surveys or medical record audits for these 'patterns of care' studies can be time- and resource-intensive. With the development of Clinical Cancer Registries in public facilities in many Area Health Services from 2006, there is now potential to conduct patterns of care studies using routinely collected data.

Clinical Cancer Registries (ClinCR) contain items about cancer diagnosis, staging, surgical procedures, adjuvant therapy, and referral and consultations with specialists that provide additional depth to data collected by the Central Cancer Registry (CCR), Admitted Patients Data Collection (APDC) and Emergency Department Data Collection (EDDC). Linkage of these datasets and feedback of patterns of treatment and outcomes that are identified may stimulate improvements in cancer service delivery.

The ClinCR is a relatively recent development and there is little experience with the use of these data for research. The purpose of this report is therefore to assess the quality, coverage and completeness of ClinCR data for use in colorectal cancer services research, and to assess the feasibility of developing surgical process and outcomes indicators that rely on ClinCR data items.

AIMS

1. To quantify the coverage of the Clinical Cancer Registry for people diagnosed with a new colorectal cancer in NSW in 2007 and 2008
2. To quantify the proportion of missing data for each item in the Clinical Cancer Registry Minimum Data Set and Colorectal Cancer Data Set extension
3. To investigate concordance between Clinical Cancer Registry and other administrative datasets for items which are duplicated across datasets
4. To investigate associations between Clinical Cancer Registry items and closely related items in other administrative datasets
5. To investigate the internal consistency of items (particularly date items) within the Clinical Cancer Registry
6. To investigate the feasibility of calculating surgical process and outcomes indicators that are dependent on Clinical Cancer Registry data

METHODS

The study sample comprised all people registered by the NSW Central Cancer Registry (CCR) with a newly diagnosed cancer of the colon, rectum or anus between 1 January 2000 and 31 December 2010.

The data used for analysis were obtained from linkage of the datasets listed below. Record linkage was conducted by the Centre for Health Record Linkage (CHeReL) using probabilistic matching of person identifiers provided by data custodians. Numbered keys generated to identify individuals across data sources were returned to custodians who released the data stripped of other identifying information.

- NSW Clinical Cancer Registries (ClinCR): Data from all six locally-managed Area Health Service (AHS) datasets which were created in 2006/2007. Information is collected only in public facilities. Items include demographic and system details, diagnostic, staging, clinical, treatment and care items.
- NSW Oncology Group dataset extension for colorectal cancer: Additional clinical data items that complement the core ClinCR dataset. These data are only collected in two AHS.
- NSW CCR: Population-based registry for cancer diagnosed or treated in NSW residents. Items for the linked dataset included demographic items, tumour details, and treatment details.
- NSW Admitted Patients Data Collection (APDC): Records for hospital separations (discharges, transfer, deaths) and demographic details from all NSW facilities.
- NSW Emergency Department Data Collection: Information on emergency department presentation within 30 days of hospital separation for individuals within the study cohort.
- NSW Registry of Births, Death and Marriages (RBDM): Information about date of death.

Data from the individual ClinCR datasets were harmonised and combined. The collated ClinCR dataset was restricted to colorectal cancer records from 2007 and 2008 (12,968 records for 5,452 persons).

APDC, RBDM and ClinCR records were merged onto the CCR cohort. CCR records were excluded if they were not the first cancer notification, were not unique first cancers, occurred before July 2000, or had inconsistent hospital or death dates indicative of an incorrect link. Data from the CCR were available to the end of 2008. Patients with an admission for a procedure of interest were flagged. The final CCR cohort included 37,593 colorectal cancer cases diagnosed between July 2000 and December 2008.

Coverage of the cohort by the ClinCR, completeness of ClinCR and dataset extension items, concordance between duplicate items across datasets, and internal consistently of ClinCR date items were examined.

RESULTS

Aim 1: Coverage of patient cohort

- 56% of colorectal cancer cases registered in the CCR in 2007 and 2008 have a ClinCR record.
- 73% of people diagnosed in 2007 and 85% of people in 2008 who had a procedure of interest in a public hospital have a ClinCR record. This proportion is over 94% for all AHS with a ClinCR in 2008.

Aim 2: Completeness of ClinCR and extension data items (for all records in the collated ClinCR dataset)

- Demographic and diagnostic items: data are available for over 95% of people for every non-optional item except postcode (64%), indigenous status (74%), and histopathological grade (85%).
- Staging items: non-optional staging items have completeness rates of 77% to 84%.
- Treatment items: treatment modality is specified in only 49% of records. All admission-related, radiotherapy and chemotherapy dates have completeness of 91% or more, except for date of chemotherapy end (87%). Specific treatment items range between 86% and 99% completeness.
- Quality of care: data are available for <60% of people for all items except psychosocial referral (69%)
- Colorectal extension items (two AHS only): most items range between 53% to 81% completeness, although completeness is lower for rectal cancer specific items.

Aims 3 and 4: Concordance with other datasets for duplicate and closely related items

- 15 ClinCR items are duplicated or are closely related to items in other datasets.
- Mismatching values make up 3% or less of data for all duplicate demographic items except for one country of birth item (25%). Greater discordance exists between staging and diagnostic items and admission-related dates (4% to 22%), though the size of the date differences are generally not large.

Aim 5: Internal consistency of ClinCR items

- 18% of ClinCR records have one or more sets of illogical dates.
- While more illogical dates arise between dates of treatment or care and date of diagnosis, there are greater median discrepancies between dates of treatment or care and date of death.

Aim 6: Feasibility of developing surgical indicators

- ClinCR or dataset extension items are required for the calculation of 12 of 21 surgical outcome and evidence-based care indicators that could feasibly be developed from the linked cohort dataset.

CONCLUSIONS

The ClinCR and dataset extension provide a range of data on cancer staging, clinical treatment and care processes that add valuable information to other routinely collected data sources. There is good coverage of public facilities, with increases from 2007 to 2008. Completeness is high for demographic and staging items, though these items are largely already collected in other datasets. Concordance with other datasets and internal consistency is also high for the majority of items and records, though some discrepancies require further exploration. While there is high completeness of treatment items, the ClinCR is limited by low rates of data for care items and difficulty determining whether records have been created for all treatment events. However, the development of surgical indicators would be greatly limited without many items from the ClinCR and dataset extension. Recommendations resulting from this report include concentrating resources on items that are unique to the ClinCR, increasing coverage to private facilities, determining accuracy of the recording of treatment events, and increasing the collection of the ClinCR and dataset extension to all AHS where they are not currently in operation.

BACKGROUND

Colorectal cancer (CRC) is the second most common registered cancer in New South Wales (NSW) and is one of the most common cancers worldwide.[1,2] There is high-quality evidence from randomised trials of specific treatment approaches that will improve patient outcomes, but these treatments must be implemented in routine clinical care to be of benefit to patients. In Australia and internationally, there is evidence that clinical care for colorectal cancer is highly variable and is often not in accordance with evidence-based clinical practice guidelines.[3-6] Within New South Wales (NSW), previous research has identified that in 2000, only 74% of patients with Dukes C colon cancer, and only around half of those with high-risk rectal cancer, were offered appropriate adjuvant chemo-radiotherapy in accordance with national evidence-based clinical practice guidelines. Older people were consistently less likely to be treated according to guidelines.[6]

Population-based studies of patients' treatments, care pathways and outcomes are essential to monitor the uptake of evidence-based practice and can identify groups who are most likely to receive sub-optimal care or experience poor outcomes. To date, such 'patterns of care' studies have relied on surveys of patients and/or clinicians and/or medical record audits to collect clinical information. These methods of data collection are resource-intensive and take a considerable length of time to assemble and follow a prospective cohort of people. With the development of the Area Health Service-based Clinical Cancer Registries, there is now potential to conduct patterns of care studies using routinely collected data. With linkage of this clinical information to other datasets, namely the Central Cancer Registry (CCR), Admitted Patients Data Collection (APDC) and Emergency Department Data Collection (EDDC) through the Centre for Health Record Linkage (CHeReL), it is theoretically possible to collate comprehensive information about patients' diagnosis, treatment and outcomes in order to stimulate improvements in cancer service delivery. However, the Clinical Cancer Registries are a relatively recent development and there is little experience with the use of these data for research purposes.

Clinical Cancer Registries (ClinCR) were established in five Area Health Services (AHS) (Sydney South West, Sydney West, North Sydney and Central Coast, South Eastern Sydney and Illawarra and North Coast Area Health Services) in 2006, with a registry in Hunter New England AHS commencing in 2007. Clinical Cancer Registries contain information about patients diagnosed or treated for cancer in public facilities within the AHS. Data items comprise a Minimum Data Set (MDS) relevant to all tumour types and cancer-specific extensions that have been developed by NSW Oncology Groups.[7] The Minimum Data Set includes items about cancer diagnosis, cancer stage, additional information about surgical procedures, adjuvant therapy, referrals and consultations with medical and radiation oncologists. A major strength of this dataset is the depth of clinical information that is collected.

For Clinical Cancer Registry data to be a useful addition to Central Cancer Registry and Admitted Patients Data Collection records in population-based record linkage research studies, the data must be available in a timely manner for a large proportion of the cohort or a representative subset (coverage), must have low levels of missing data (completeness) and must be accurate.

The purpose of the analyses contained in this report is to assess the quality and completeness of Clinical Cancer Registry data for use in colorectal cancer services research, and specifically to assess the feasibility of developing surgical process and outcomes indicators that rely on Clinical Cancer Registry data items for the numerator, denominator or for risk adjustment. The report pertains only to colorectal cancer as it is part of a larger data linkage study to develop and test composite indicators of colorectal cancer care using population-based administrative datasets. Similar assessments should be undertaken for other cancer sites.

AIMS

1. To quantify the coverage of the Clinical Cancer Registry for people diagnosed with a new colorectal cancer in NSW in 2007 and 2008
2. To quantify the proportion of missing data for each item in the Minimum Data Set and Colorectal Cancer Data Set extension
3. To investigate concordance between Clinical Cancer Registry and other administrative datasets for items which are duplicated across datasets
4. To investigate associations between Clinical Cancer Registry items and closely related items in other administrative datasets
5. To investigate the internal consistency of items (particularly date items) within the Clinical Cancer Registry
6. To investigate the feasibility of calculating surgical process and outcomes indicators that are dependent on Clinical Cancer Registry data

METHODS

Study cohort

The study sample comprised all people registered by the NSW Central Cancer Registry (CCR) with a newly diagnosed cancer of the colon, rectum or anus (ICD-10-AM codes C18- C21) between 1 January 2000 and 31 December 2010.

Clinical Cancer Registry (ClinCR) data

Six AHS-based Clinical Cancer Registries

Clinical cancer registries (ClinCR) were established in five Area Health Services (AHS) (North Coast, Northern Sydney Central Coast, South Eastern Sydney Illawarra, Sydney South West, and Sydney West AHS) in 2006, and in Hunter New England AHS in 2007. Clinical Cancer Registries were not established in Greater Southern AHS or Greater Western AHS. Each ClinCR is managed locally, with data collection, data entry, quality control and data governance the responsibility of the AHS. It should be noted that the ClinCR covers only patients treated within public facilities within the AHS. Patients treated in private facilities are not included.

Data items contained in the ClinCR include:

- Demographic details: sex, age, country of birth, postcode, indigenous status
- System details: AMO registration number, facility code
- Diagnostic and staging items: date of diagnosis, primary site of cancer, best basis for diagnosis, histopathological grade, morphology, degree of spread, TNM staging
- Clinical details: performance status at diagnosis, date of death, cause of death
- Treatment items: date of admission, date of discharge, procedure codes, date of radiotherapy start, date of radiotherapy end, radiotherapy type, dose and fractions, date of systemic therapy start, date of systemic therapy end, systemic protocol, number of cycles, date of referral to cancer specialist, date of consultation with specialist, date of decision to treat, date of clinical trial enrolment, date of MDT, date of referral to palliative care, psycho-social referral to (type)

The NSW Oncology Group for colorectal cancer also identified a dataset extension to complement the core clinical cancer dataset. The colorectal dataset extension aims to include the following 12 items:

- Presentation (screening/symptoms/emergency), method of surgery, level of rectal cancer, residual tumour status, radial resection margin, lymphovascular invasion, number of nodes examined, number of nodes involved, date of resection, mismatch repair deficiency (MMRD), site of recurrence, date of recurrence

Creation of collated ClinCR dataset

Access to the clinical cancer registries was provided for all six AHS: Sydney South West AHS (SSW), North Coast AHS (NC), Northern Sydney Central Coast AHS (NSCC), South Eastern Sydney Illawarra AHS (SESI), Sydney West AHS (SW), and Hunter New England AHS (HNE). Extension item data were also available for two AHS. To protect their identity, areas have been labeled AHS A to AHS F for this report (in no particular order).

Many data items were common across all ClinCRs, however some were unique to individual Area datasets and many were recorded in different ways. Data were harmonised to best resemble the current data dictionary and to enable stacking (combining) of the datasets. This process included standardising data item names, cleaning erroneous data, standardising default codes, transforming data recorded as text strings to standard codes, determining meaning of Area-specific coding not recorded in the minimum dataset dictionary and reassigning codes, standardising items where different coding systems had been used (e.g. systemic protocol). Records containing no data besides PPN were also deleted.

The different area datasets originally recorded procedure codes in different formats. Procedure items were collapsed so that all procedures for an admission were recorded in one row (up to 27). Duplicate records providing no extra information to other records were deleted. Records were then stacked (combined) to create one ClinCR dataset.

While data from some AHS registries were available for those diagnosed in and before 2006, greater levels of data were recorded in all registries after 2006. Information was also available for persons who had subsequent cancers of other sites in some datasets. For this analysis, the ClinCR dataset was restricted to records with a date of diagnosis in 2007 and 2008 for which primary cancer site was recorded as colon, rectum or anus. Table 1 shows the flow of records and persons in creating the collated ClinCR. 12,968 records for 5,452 persons were included in the final dataset.

Table 1: Flow of records in collated ClinCR creation

	CHeReL linked records (persons)	Received records with data (persons)	Records after procedure collapse (persons)	Records after deleting duplicates (persons)	Restricted dataset records (persons)*
AHS A	1,672 (585)	1,566 (503)	1,561 (503)	1,561 (503)	1,517 (487)
AHS B	2,451 (953)	2,554 (953)	2,554 (953)	2,553 (953)	1,629 (586)
AHS C	5,734 (2,390)	5,823 (2,183)	5,765 (2,183)	5,752 (2,183)	2,625 (1,002)
AHS D	4,011 (2,082)	4,011 (2,082)	4,011 (2,082)	4,011 (2,082)	2,367 (1,265)
AHS E	5,250 (4,012)	9,873 (4,011)	6,514 (4,011)	6,513 (4,011)	2,245 (1,462)
AHS F	3,999 (1,416)	14,246 (1,416)	4,291 (1,416)	4,275 (1,416)	2,585 (873)
Total	23,118 (10,986)	38,073 (10,723)	24,696 (10,723)	24,665 (10,723)	12,968 (5,452)

* Restricted to colorectal cancer diagnosis in 2007 and 2008

Other data sets and data items available in colorectal linked dataset

NSW Central Cancer Register (CCR)

The NSW CCR was established in 1971 as a population-based registry for all cases of cancer diagnosed or treated in residents of NSW. Notification of cancer to the Registry is mandatory under the NSW Public Health Act 1991. The data items contained in this data collection that will were included in the linked dataset for the project include:

- Demographic details (year of birth, age at diagnosis, sex, country of birth, indigenous status, postcode of residence, area health service (AHS) of residence, date of death, cause of death)
- Tumour details (date of diagnosis, primary site, basis of diagnosis, morphology, degree of spread)
- Treatment details (episode type, episode date, date of admission, date of separation, hospital code, AHS of hospital, hospital sector (public/private), principal procedure).

NSW Admitted Patients Data Collection (APDC)

The Admitted Patient Data Collection (APDC) is a population-based data collection that includes records for all hospital separations (discharges, transfers and deaths) from all NSW public and private hospitals and day procedure centres. Data items from this collection that were included in the linked dataset for the project include:

- Demographic details (eg age, sex, country of birth, indigenous status, language spoken at home, marital status, postcode of residence)
- Separation details (date of admission, date of separation, source of referral, discharge status and diagnosis and procedure codes using ICD-10-AM coding, including coding of significant comorbidities and complications).

Validation studies have found approximately 95% agreement between principal diagnoses recorded in hospital inpatient statistics and corresponding clinical records.[8] Data on procedures (which have financial implications for an organisation) are accurately recorded in almost all cases.[8] ICD diagnosis codes were used to calculate a Charlson co-morbidity index.[9]

NSW Emergency Department Data Collection

Data in this dataset are derived from the Emergency Department Information System. Although the NSW Department of Health has highlighted that, over the longer term, lost linkages will occur when using this dataset (estimated at approximately 10% per year when hospital code and medical record

number are used for linkage), linkage of events that are close in time (within a few weeks) is considered to be reliable.[10] For this project, linkage was restricted to a period of 30 days from hospital separation for individuals within the study cohort, so that rates emergency department presentation within 30 days can be assessed. Information is available from 1 Jan 2005.

NSW Registry of Births, Deaths and Marriages (RBDM)

Information about date of death was obtained from this source.

Record linkage

Record linkage was conducted by the Centre for Health Record Linkage (CHeReL) using probabilistic matching. Probabilistic matching conducted by the CHeReL involves data custodians providing person identifiers without any health information to the CHeReL. The CHeReL matches these identifiers across data sources, and generates a system of numbered keys to identify individuals. After providing data custodians with these keys, the custodians can release health information to researchers, identified only by the numbered keys and stripped of all other identifying information. At no time do researchers have access to identifying information, nor does the CHeReL have access to health information. This approach to preserving privacy and data governance has been strongly supported by organisations that are custodians of health records, human research ethics organisations, researchers and the community.

The CHeReL's probabilistic linkage procedures are designed to achieve a false positive rate around 5/1,000. A random sample of 1000 Person IDs from the linkage was selected and reviewed by the CHeReL, and rate of the false positives was 3/1,000.

Statistical analysis

Analysis subset

The set of data used for analysis was obtained from linkage between CCR, NSW APDC and ClinCR data. Cancer notifications from the CCR were included for the first notification, with only unique first cancers included. Data from the CCR were available to the end of 2008. We selected cancers from July 2000 onwards, to be consistent with APDC records which are collected by financial year. Included cancer notifications were merged onto the APDC to identify hospitalisations for selected cancers. Consistency of dates of admission, procedure, separation and death were checked in the APDC, with inconsistent records ignored. Episodes of care were created, to allow for transfers between hospitals within an admission. Admissions with a procedure of interest occurring in or after the month of cancer diagnosis were selected as index admissions. Procedures of interest, summarised in Appendix 1, were derived from work conducted with the MERU team at the Cancer Institute NSW, and supplemented by procedures used in Hall et al (2005).[11]

Coverage

Coverage was examined in several ways to inform various potential uses of the data. The proportion of the cohort with a ClinCR record was calculated for the total cohort (population coverage for those diagnosed since 2000) and for those diagnosed in 2007 and 2008 (when the ClinCR was in operation). The proportion of the cohort diagnosed in 2007 and 2008 with a procedure of interest in a public hospital who have a ClinCR record was also determined, and the results tabulated by AHS.

Completeness

People in the ClinCR have more than one record if they received multiple occasions of treatment and/or multiple treatment types. Levels of missing data for each item in the collated ClinCR and dataset extension were therefore examined in 3 ways: 1) the number and proportion of records with data recorded for the item, 2) the number and proportion of persons who have any data recorded for the item, and 3) the number and proportion of persons for whom all records have data for the item.

Concordance

Concordance between items in the CCR/APDC/RBDM and duplicate items in the ClinCR was examined for those in the cohort with a ClinCR record. The proportion of people with the same data values across duplicate items was calculated. As many of the data were skewed to outliers, median and interquartile range (IQR) were used in reporting the differences between discordant dates.

Internal consistency

The number of records and persons in the collated ClinCR for whom dates appear illogical was calculated (e.g. date of radiotherapy before date of diagnosis), as well as the median difference (IQR) between illogical dates.

RESULTS

Description of cohort

We selected 37,593 first unique cancer notifications between July 2000 and December 2008 from a total of 40,328 cancer notifications received from the CCR. A total of 37,050 people had data recorded in the APDC, and 437,695 records from 37,049 people had consistent dates of admission, procedure, separation and death. There were 28,360 patients with an admission containing a procedure of interest in or after the month previous to cancer notification.

Linked dataset

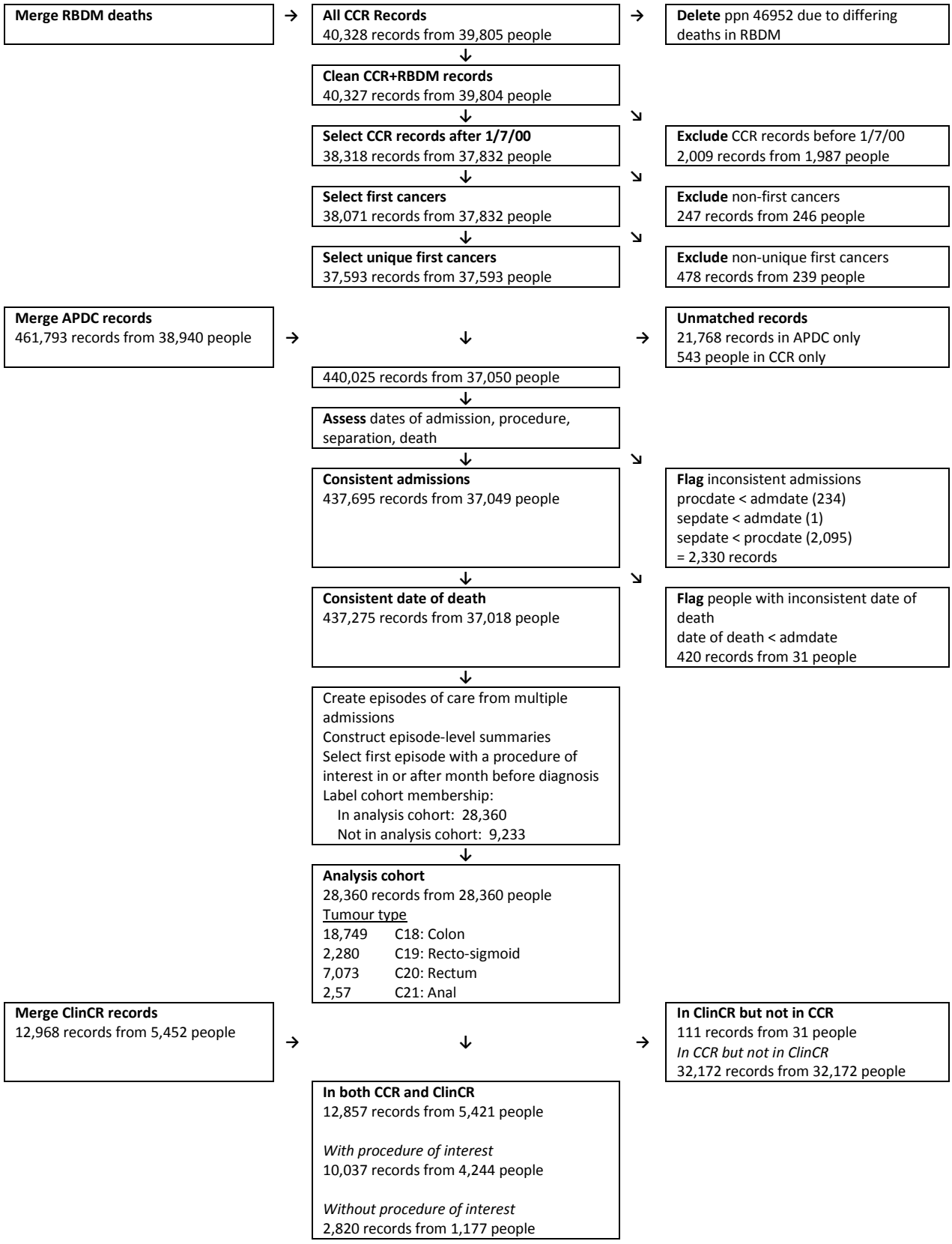
The characteristics of the people in the ClinCR dataset and those in the study cohort are compared in Table 2. A flowchart for the creation of the linked dataset is shown in Figure 1.

Table 2: Characteristics of persons in the ClinCR and study cohort

	ClinCR n persons (%)	**Study cohort* n persons (%)**
Sex		
Male	3,033 (55.6)	20,441 (54.4)
Female	2,418 (44.4)	17,152 (45.6)
Age at diagnosis		
0-49	408 (7.5)	2,567 (6.8)
50-59	791 (14.5)	5,547 (14.8)
60-69	1,398 (25.6)	9,696 (25.8)
70-79	1,606 (29.5)	11,546 (30.7)
80+	1,248 (22.9)	8,237 (21.9)
Country of birth		
Australia	3,519 (64.6)	26,109 (69.5)
Other country	1,883 (34.5)	9,992 (26.6)
Missing	50 (0.9)	1,492 (4.0)
Primary site of cancer		
Colon	3,355 (61.5)	24,172 (64.3)
Rectosigmoid junction	498 (9.1)	2,980 (7.9)
Rectum	1,459 (26.8)	9,642 (25.6)
Anus	140 (2.6)	799 (2.1)
Degree of spread		
Localised	1,849 (33.9)	12,702 (33.8)
Regional spread	2,147 (39.4)	14,718 (39.2)
Distant metastases	1,173 (21.5)	6,253 (16.6)
Missing	283 (5.2)	3,920 (10.4)
TOTAL	5,452 (100.0)	37,593 (100.0)

* Based on CCR items

Figure 1: Flowchart of linked dataset creation



1. Data availability

The study cohort comprised people with an incident colorectal or anal cancer diagnosed from 1 January 2000 onwards. At the time of this analysis, data from the CCR were available to 31 December 2008. Data from the APDC were available to the middle of 2010, and the RBDM to the end of 2010. ClinCR data for 2007 and 2008 were included in the analysis.

2. Coverage of patient cohort

99.4% of people in the ClinCR linked to a record in the cohort dataset (5,421 of 5,452 persons). While only 14.4% of people in the total cohort have a ClinCR record, 55.9% of people diagnosed in 2007 and 2008 have a ClinCR record.

Table 3: Proportion of cohort with a ClinCR record

	Persons in cohort	Persons in cohort with ClinCR record, n (%)
Pre 2007	28,136	130*
2007	4,803	2,442 (50.8)
2008	4,654	2,849 (61.2)
Total 2007/2008	9,457	5,291 (55.9)
Total	37,593	5,421 (14.4)

* Year of diagnosis >=2007 in ClinCR but discordant with linked record

Proportion of cohort with a procedure of interest in a public hospital (PH) who have a ClinCR record

78.5% of people diagnosed in 2007 and 2008 who had a procedure of interest in a public hospital have a ClinCR record. The coverage of the ClinCR for procedure of interest in a public hospital increased from 72.7% for those diagnosed in 2007 to 84.7% in 2008.

Table 4: Proportion of cohort with a procedure of interest in a PH who have a ClinCR record by year

	Persons in cohort	Persons in cohort with a procedure of interest in a PH, n (%)	Persons in cohort with a procedure of interest in a PH with a ClinCR record, n (%)
Pre 2007	28,136	13,042 (46.4)	81*
2007	4,803	2,185 (45.5)	1,589 (72.7)
2008	4,654	2,061 (44.3)	1,746 (84.7)
Total 2007/2008	9,457	4,246 (44.9)	3,335 (78.5)

* Year of diagnosis >=2007 in ClinCR but discordant with linked record

By 2008, all AHS had a ClinCR in full operation. The proportion of people who had a procedure of interest in a PH who have a ClinCR record is over 94% for all AHS in 2008.

Table 5: Proportion of cohort with a procedure of interest in a PH who have a ClinCR record by AHS

	Persons in cohort with a procedure of interest in a PH, 2008 only	Persons in cohort with a procedure of interest in a PH with ClinCR record, n (%)
AHS A	283	278 (98.2)
AHS B	207	196 (94.7)
AHS C	226	218 (96.5)
AHS D	370	360 (97.3)
AHS E	482	463 (96.1)
AHS F	222	217 (97.7)
Total	1,790	1,732 (96.8)

3. Completeness of data items

This section aims to determine the completeness of ClinCR data in terms of its usefulness for other analyses. For the purposes of this report, data coded as "unknown", "not stated", "uncertain" or as default dates are considered missing. Codes for data items are available in Appendix 2 and Appendix 3.

People in the ClinCR have more than one record if they received multiple occasions of treatment and/or multiple treatment types. Missingness is reported below both by record and by person, but the denominator varies depending on whether the record can reasonably be expected to have data for a certain item. For example, a record relating to radiotherapy would not be expected to have information on chemotherapy items or hospital admissions, therefore the proportion of available data is based on those records relating to radiotherapy alone.

4.1% of persons in the ClinCR dataset are treated in more than one AHS (see table 7 below). To more accurately determine the proportion of persons for whom at least some data are available for a particular item ("persons with any data"), the first AHS where a person receives treatment is credited with the non-missing data. All person data is reported by first AHS of presentation, unless otherwise stated. A breakdown of completeness of each item by AHS is available in Appendix 4 and Appendix 5.

Table 6: Records and persons in the ClinCR by AHS

	n records (%)	n persons (%)
AHS A	1,517 (11.7)	473 (8.7)
AHS B	1,629 (12.6)	577 (10.6)
AHS C	2,625 (20.2)	954 (17.5)
AHS D	2,367 (18.3)	1,232 (22.6)
AHS E	2,245 (17.3)	1,395 (25.6)
AHS F	2,585 (19.9)	821 (15.1)
Total ClinCR	12,968 (100.0)	5,452 (100.0)

Table 7: Persons in multiple AHS in the ClinCR

	n persons (%)
Persons in one AHS	5,230 (95.9)
Persons in two AHS	221 (4.1)
Persons in three AHS	1 (0.0)
Total	5,452 (100.0)

Demographic items

All records were included in the denominator for demographic items.

Table 8: Completeness of demographic items in the ClinCR

	Records with data, n (%)	Persons with *any* data, n (%)	Persons with *all* data, n (%)
Sex	12,966 (100.0)	5,450 (100.0)	5,450 (100.0)
Year of birth	12,967 (100.0)	5,451 (100.0)	5,451 (100.0)
Country of birth	12,876 (99.3)	5,402 (99.1)	5,392 (98.9)
Postcode	8,015 (61.8)	3,469 (63.6)	3,329 (61.6)
Indigenous status	10,630 (82.0)	4,058 (74.4)	3,932 (72.1)
AMO registration number	12,732 (98.2)	5,299 (97.2)	5,230 (95.9)
Facility code	12,921 (99.6)	5,411 (99.2)	5,405 (99.1)

Diagnostic items

All records were included in the denominator for diagnostic items. ClinCR data dictionary item "laterality of primary cancer" was not recorded by any AHS.

Table 9: Completeness of diagnostic items in the ClinCR

	Records with data, n (%)	Persons with *any* data, n (%)	Persons with *all* data, n (%)
Date of diagnosis of cancer	12,968 (100.0)	5,452 (100.0)	5,452 (100.0)
Primary site of cancer	12,968 (100.0)	5,452 (100.0)	5,452 (100.0)
Best basis for primary cancer diagnosis	12,949 (99.9)	5,446 (99.9)	5,436 (99.7)
Histopathological grade	10,884 (83.9)	4,654 (85.4)	4,543 (83.3)
Morphology of cancer	12,961 (99.9)	5,450 (100.0)	5,450 (100.0)
Morphology (ICD) version*	1,623 (12.5)	585 (10.7)	558 (10.2)
Degree of spread of cancer	12,424 (95.8)	5,214 (95.6)	5,146 (94.4)

*** N**ot in ClinCR data dictionary and not requested by researchers; reported by one AHS only

Staging items

All records were included in the denominator for staging items.

Table 10: Completeness of staging items in the ClinCR

	Records with data, n (%)	Persons with *any* data, n (%)	Persons with *all* data, n (%)
Cancer staging - T stage code	10,315 (79.5)	4352 (79.8)	4,238 (77.7)
Cancer staging - N stage code	10,198 (78.6)	4,298 (78.8)	4,194 (76.9)
Cancer staging - M stage code	10,040 (77.4)	4,213 (77.3)	4,103 (75.3)
Staging basis	10,678 (82.3)	4,597 (84.3)	4,495 (82.4)
Other staging systems and classifications[1]	4,396 (33.9)	2,374 (43.5)	2,246 (41.2)
Other stage groupings[2]	4,348 (33.5)	2,335 (42.8)	2,207 (40.5)
TNM stage[3]	7,091 (54.7)	3,384 (62.1)	3,240 (59.4)
TNM edition number[3]	3,845 (29.6)	1,332 (24.4)	1,246 (22.9)
Best stage[3]	2,184 (16.8)	1405 (25.8)	1,294 (23.7)

[1] Optional item; [2] Dependent on above item

[3] Not in ClinCR data dictionary and not requested by researchers

Treatment items

Treatment items in the ClinCR include those related to admissions, radiotherapy, and chemotherapy. ClinCR data dictionary item "treatment modality" (below), which indicates type of treatment for each record, was available in 3 of the 6 AHS.

	Records with data, n (%)	**Persons with *any* data, n (%)**	**Persons with *all* data, n (%)**
Treatment modality	6,347 (48.9)	2,760 (50.6)	2,695 (49.4)

Admissions

Records were flagged as admission records if they had a value for any of the admission items (date of admission, date of discharge, treatment procedure) or the item "treatment modality" indicated admission. These records became the denominator for admission-related items. The total number and proportion of admission records in the ClinCR dataset are presented below, followed by a breakdown of completeness by item.

	n records (%)	**n persons (%)**
Total admissions	8,794 (67.8)	4,820 (88.4)

Table 11: Completeness of treatment items in the ClinCR (admission-related)

	Records with data, n (%)	**Persons with *any* data, n (%)**	**Persons with *all* data, n (%)**
Date of admission	8,648 (98.3)	4,761 (98.8)	4,683 (97.2)
Date of discharge	8,610 (97.9)	4,748 (98.5)	4,651 (96.5)
Procedure (ICD) version[1]	4,422 (50.3)	2,049 (42.5)	1,705 (35.4)
Treatment procedure for cancer (1)[2]	8,232 (93.6)	4,787 (99.3)	4,407 (91.4)

[1] Not in ClinCR data dictionary and not requested by researchers

[2] There are 27 procedure items; this item reflects how many people have at least one procedure code recorded

Radiotherapy

Records were flagged as radiotherapy records if they had a value for any of the radiotherapy items or the item "treatment modality" indicated radiotherapy. These records became the denominator for radiotherapy-related items. The total number and proportion of radiotherapy records in the ClinCR dataset is presented below, followed by a breakdown of completeness by item.

	n records (%)	**n persons (%)**
Total radiotherapy	812 (6.3)	762 (14.0)

Table 12: Completeness of treatment items in the ClinCR (radiotherapy-related)

	Records with data, n (%)	**Persons with *any* data, n (%)**	**Persons with *all* data, n (%)**
Date of radiotherapy start	760 (93.6)	719 (94.4)	711 (93.3)
Date of radiotherapy end	739 (91.0)	705 (92.5)	699 (91.7)
Radiotherapy treatment type	762 (93.8)	737 (96.7)	725 (95.1)
Received radiation dose	700 (86.2)	682 (89.5)	668 (87.7)
Radiation fractions	704 (86.7)	685 (89.9)	672 (88.2)

Chemotherapy

Records were flagged as chemotherapy records if they had a value for any of the chemotherapy items or the item "treatment modality" indicated chemotherapy. These records became the denominator for chemotherapy-related items. The total number and proportion of chemotherapy records in the ClinCR is presented below, followed by a breakdown of completeness by item.

	n records (%)	**n persons (%)**
Total chemotherapy	3,290 (25.4)	2,469 (45.3)

Table 13: Completeness of treatment items in the ClinCR (chemotherapy-related)

	Records with data, n (%)	**Persons with *any* data, n (%)**	**Persons with *all* data, n (%)**
Date of systemic therapy start	3,102 (94.3)	2,331 (94.4)	2,289 (92.7)
Date of systemic therapy end	2,869 (87.2)	2,196 (88.9)	2,083 (84.4)
Systemic therapy protocol name	3,275 (99.5)	2,457 (99.5)	2,454 (99.4)
Number of cycles	2,921 (88.8)	2,219 (89.9)	2,131 (86.3)

Quality of care indicators

All records were part of the denominator for quality of care indicators. ClinCR data dictionary item "Hospital Anxiety and Depression Scale" was not recorded by any AHS.

Table 14: Completeness of quality of care items in the ClinCR

	Records with data, n (%)	**Persons with *any* data, n (%)**	**Persons with *all* data, n (%)**
Date of referral to cancer specialist	3,708 (28.6)	2,284 (41.9)	709 (13.0)
Date of consultation with cancer specialist	5,901 (45.5)	3,209 (58.9)	1,648 (30.2)
Date of decision to treat	844 (6.5)	618 (11.3)	143 (2.6)
Date of first clinical trials enrolment	451 (3.5)	132 (2.4)	107 (2.0)
Date of first multidisciplinary team consultation (MDT)	3,707 (28.6)	1,254 (23.0)	1,184 (21.7)
Date of referral to palliative care	3,657 (28.2)	1,157 (21.2)	1,082 (19.8)
Performance status at diagnosis	3,506 (27.0)	1,386 (25.4)	1,283 (23.5)
Psycho-social referral to	9,657 (74.5)	3,786 (69.4)	3,663 (67.2)
Date of death	4,174 (32.2)	1,305 (23.9)	1,240 (22.7)
Cause of death	656 (5.1)	376 (6.9)	344 (6.3)

NSW Oncology Group Colorectal Extension

Colorectal extension data were available for two AHS only, labeled AHS X and AHS Y below. Any record with a value for any extension item became part of the denominator for calculating completeness of extension items. Extension data dictionary items "date of resection", "site of first recurrence", and "date of first recurrence" were not provided by either AHS.

	n records (% of records in ahs)	**n persons (% of persons in ahs)**
AHS X	[redacted] (92.7)	[redacted] (92.4)
AHS Y	[redacted] (54.3)	[redacted] (52.1)
Total extension	3,413 (26.3)	1,908 (35.0)

[*] Blacked out to preserve AHS identity

Table 15: Completeness of NSW Oncology Group colorectal extension items

	Records with data, n (%)	Persons with *any* data, n (%)	Persons with *all* data, n (%)
Presentation	2,641 (77.4)	1,384 (72.5)	1,371 (71.9)
Method of surgery	1,817 (53.2)	1027 (53.8)	1,023 (53.6)
Level of rectal cancer	434 (12.7)	186 (9.7)	181 (9.5)
Residual tumour status	2,610 (76.5)	1,452 (76.1)	1,445 (75.7)
Radial resection margin	1,181 (34.6)	664 (34.8)	656 (34.4)
Lymphovascular invasion	2,637 (77.3)	1,474 (77.3)	1,469 (77.0)
Number of lymph nodes examined	2,751 (80.6)	1,534 (80.4)	1,529 (80.1)
Number of lymph nodes involved	2,760 (80.9)	1,539 (80.7)	1,532 (80.3)
Mismatch repair deficiency (MMRD)	2,699 (79.1)	1,501 (78.7)	1,496 (78.4)
Perineural invasion*	2,722 (79.8)	1,509 (79.1)	1,503 (78.8)

* Not in extension data dictionary and not requested by researchers; collected by both AHS

One AHS collected a number of additional colorectal items, however these are not presented in this report.

4. Concordance with other datasets for duplicate items

Data for items in the CCR, APDC and RBDM were compared to data in duplicate items in the ClinCR. While there is only one record per person in the cohort dataset, there are multiple records per person in the ClinCR dataset. Records were considered to be concordant if data from an item in the CCR/ APDC/ RBDM matched at least one ClinCR record for the same person, or if both item values were missing. The difference between mismatching dates were calculated based on the closest ClinCR date. Dates within two days of date of death and one month of date of diagnosis were considered concordant.

Table 16: Concordance of items between the ClinCR and other datasets

	Discordant (%), mismatch	Discordant (%), one missing	Concordance (%)	Days difference, median (IQR)
Sex (CCR)	0.1	0.0	99.9	
Sex (APDC)	0.1	0.0	99.9	
Age at diagnosis (CCR)	2.1	0.0	97.8	4 (3) years
Indigenous status (CCR)	0.4	25.8	73.7	
Indigenous status (APDC)	0.3	25.2	74.5	
Country of birth[1] (CCR)	1.3	3.7	95.0	
Country of birth[1] (APDC)	25.3	1.9	72.8	
Postcode (CCR)	3.1	36.5	60.5	
Postcode (APDC)	2.3	38.1	59.6	
Area health service (CCR)	2.4	21.5	76.1	
Cancer site[2] (CCR)	10.6	0.0	89.4	
Morphology (CCR)	19.1	0.0	80.9	
Degree of spread (CCR)	22.2	9.1	68.7	
Date of diagnosis (CCR)	4.3	0.0	95.7	3 (2) months[3]
Date of admission (APDC)	11.2	11.0	77.7	2 (7)
Date of discharge (APDC)	12.9	11.3	75.7	10 (18)
Date of death (CCR)	0.2	16.6	83.1	10 (21)[4]
Date of death (RBDM)	0.5	16.0	83.5	10 (14)[4]

[1] Matching by country region (first two digits of country code)

[2] Matching by sub-categories of site, colon (C18), rectosigmoid (C19), rectum (C20), anus (C21)

[3] For date of diagnosis > 2006

[4] For matching hospital and procedure of interest or matching hospital and ClinCR date within CCR date

Other closely related variables between datasets include best basis of diagnosis (CCR) and best basis for primary cancer diagnosis (ClinCR), and degree of spread (CCR) and TNM staging (ClinCR).

5. Internal consistency of ClinCR items

The number of records and persons with illogical dates are presented in the table below. The size of the discrepancies are reported as median values in days. Illogical dates were flagged according to ClinCR data dictionary guidelines. Dates within two days of date of death were not considered illogical.

Table 17: Internal consistency of date items in the ClinCR

	< date of diagnosis, records (persons)	Days difference, median (IQR)	> date of death, records (persons)	Days difference, median (IQR)
Date of admission	1,495 (1,449)	2 (5)	4(2)	507 (545)
Date of discharge	110 (102)	14 (24)	5 (3)	245 (554)
Date of radiotherapy start	3 (3)	94 (240)		
Date of radiotherapy end	2 (2)	100 (194)		
Date of systemic therapy start	19 (19)	34 (105)	1 (1)	223 (0)
Date of systemic therapy end	18 (18)	68 (108)	13 (10)	75 (41)
Date of referral to specialist			2 (2)	174 (116)
Date of specialist consultation	1,074 (908)	10 (19)	3 (3)	116 (228)
Date of decision to treat	9 (8)	9 (7)		
Date of MDT	94 (41)	13 (46)	4(4)	17 (304)
Date of referral to palliative care	33 (17)	14 (28)	1 (1)	7 (0)
	> date of systemic therapy end		**< date of specialist consultation**	
Date of systemic therapy start	10 (10)	213 (49)		
Date of decision to treat			89 (63)	7 (6)

Some of the discrepancies above will be a result of incorrect recording of data values. Others may be illogical but still accurate. The majority of people in the ClinCR datasets (72.7%) did not have any illogical dates recorded (see table below).

Table 18: Records and persons with illogical date flags in the ClinCR

	n records (% of ClinCR)	n persons (% of ClinCR)
1 flag	1,690 (13.0)	1,486 (27.3)
2 flags	583 (4.5)	596 (10.9)
3 flags	31 (0.2)	31 (0.6)
4 flags	10 (0.1)	9 (0.2)
Total	2,314 (17.8)	1,998 (36.6)
Total ClinCR	12,968 (100.0)	5,452 (100.0)

6. Feasibility of developing surgical indicators using Clinical Cancer Registry data

Indicator	Feasible	Numerator	Needs ClinCR/DE?	Available n (%)	Denominator	Needs ClinCR/DE?	Available n
Short-term surgical outcome measures							
30-day mortality	Yes	Date of death (from CCR or RBDM)	No	100%	All patients receiving surgery	No	28,360
Proportion of patients presenting to emergency department within 28 days of discharge from hospital	No	Data from Emergency Department Data Collection for patients presenting to selected public EDs	No	100%	Patients resident near a hospital that supplies EDDC data	No	?
Proportion of patients with an ~~unplanned~~ readmission to hospital within 28 days of discharge	Yes	Date of next admission within 28 days	No	100%	All patients receiving surgery	No	28,360
Proportion of patients with an unplanned return to theatre	No	Return to theatre variable on APDC is not consistently completed across hospitals			All patients receiving surgery	No	28,360
Proportion of patients developing an anastomotic leak	No	'Complications' may be able to be detected			All patients receiving surgery	No	28,360
Proportion of patients developing a wound infection	Yes	Subsequent admissions with wound infection codes (to be defined)	No	100%	All patients receiving surgery	No	28,360
Proportion of patients developing cardiopulmonary complication	Yes	Subsequent admissions with cardio-pulmonary codes (to be defined)	No	100%	All patients receiving surgery	No	28,360
Proportion of patients developing a deep vein thrombosis	Yes	Subsequent admissions with DVT codes (to be defined)	No	100%	All patients receiving surgery	No	28,360
Proportion of patients with prolonged length of stay for the relevant Diagnostic Related Group (DRG)	Yes	Episode length of stay greater than DRG length of stay	No	100%	All patients receiving surgery	No	28,360
Proportion of patients with length of stay greater than 21 days	Yes	Episode length of stay greater than 21 days	No	100%	All patients receiving surgery	No	28,360
Longer term outcome measures							
1-year and 3-year cancer-specific/all cause survival proportions	Yes	Date of death (from CCR or RBDM)	No	100%	All patients receiving surgery	No	28,360

Indicator	Feasible	Numerator	Needs ClinCR/DE?	Available n (%)	Denominator	Needs ClinCR/DE?	Available n
Proportion of patients undergoing resection for rectal cancer who have a permanent stoma formed	Yes	Rectal cancer patients undergoing abdominoperineal excision (code: 32039-00)	No	100%	All rectal cancer patients receiving surgery	No	C19: 2,280 C20: 7,073
Evidence-based or care process measures							
Proportion of patients discussed at a multi-disciplinary team (MDT) meeting	Yes: if missing MDT date is an indication of no MDT	Date of MDT	Yes	1,035 (24.4%)	All patients with ClinCR records that link to CCR-APDC records	Yes: ClinCR	4,244
Proportion of patients with a pre-operative CT scan	No	Not measured in ClinCR / DE			All patients with ClinCR records that link to CCR-APDC records	Yes: ClinCR	4,244
Proportion of patients with lower rectal cancer with a pre-operative pelvic MRI scan	No	Not measured in ClinCR / DE			All patients with ClinCR records that link to CCR-APDC records	Yes: ClinCR	4,244
Proportion of patients with documented cancer stage	Yes	Extent of disease	No	100% [Note: 5.3% with unknown stage(=9)]	All patients receiving surgery	No	28,360
		Summary TNM stage Constructed TNM Summary or constructed TNM	Yes Yes Yes	2,646 (62.4%) 3,082 (72.6%) 3,496 (82.4%)	All patients with ClinCR records that link to CCR-APDC records	Yes	4,244
Proportion of patients having any surgical intervention (including stent)	No	Only have admissions to NSW hospitals. There are low rates of surgery for LHDs near ACT and other NSW borders, indicating cross-border hospital data are needed			All patients in CCR	No	37,593
Proportion of patients with rectal cancer seen pre-operatively by a stoma therapist	No	Not measured in ClinCR / DE					

Indicator	Feasible	Numerator	Needs ClinCR/DE?	Available n (%)	Denominator	Needs ClinCR/DE?	Available n
Proportion of patients having minimally-invasive surgery	No. Can obtain proportion of surgery that was minimally invasive	Procedures performed laparoscopically, or with stent only	No	100%	All patients receiving surgery	No	28,360
Proportion of patients with apical lymph node status recorded	No	Not measured in ClinCR / DE					
Proportion of patients with circumferential margin status recorded	Yes	Patients with radial margin status recorded	Yes: DE	621 (29.7%)	Patients with DE	Yes: DE	2,088
Proportion of patients with extramural vascular invasion status recorded	Yes	Patients with lymphovascular invasion recorded	Yes: DE	1,353 (64.8%)	Patients with DE	Yes: DE	2,088
Proportion of patients with pathological stage recorded	Yes	(T-stage, N-stage, M-stage all present OR TNM-stage present OR other stage present) AND staging basis is recorded	Yes: DE	2,386 (56.2%)	All patients with ClinCR records that link to CCR-APDC records	Yes: ClinCR	4,244
Proportion of patients receiving pre-operative thrombo-embolism prophylaxis	No	Not measured in ClinCR / DE					
Proportion of patients receiving pre-operative antibiotic prophylaxis	No	Not measured in ClinCR / DE					
Proportion of patients having Total Mesorectal Excision (TME) for lower rectal cancer	No	TME not recorded in any dataset					
Proportion of patients receiving adjuvant chemotherapy for lymph node positive colon cancer within 8 weeks of surgery	Yes	Patients with a first date of chemotherapy received	Yes: ClinCR	751 (67.0%)	Colon cancer (CCR C18) Lymph node positive (ClinCR: N-stage=1,2,3) Surgery (APDC)	Yes: ClinCR	1,121

Indicator	Feasible	Numerator	Needs ClinCR/DE?	Available n (%)	Denominator	Needs ClinCR/DE?	Available n
Proportion of patients receiving neoadjuvant/adjuvant radiotherapy for locally advanced, lower rectal cancer	Yes	All patients, assuming that no date of radiotherapy means no radiotherapy was received	Yes: ClinCR	108 (100%)	Rectal cancer (CCR C20) Locally advanced (T-stage=2,3) [Lower rectal cancer, ≤10cm from anal verge] OR [location of rectal cancer=low (SSW)]	No Yes: ClinCR Yes: DE	108
Proportion of patients receiving colonic pouch reconstruction after low anterior resection for T3 rectal cancer less than 10cm from anal verge	Yes	Patients receiving subsequent colonic pouch reconstruction (code 32029-00: Construction of colonic reservoir)	No	100%	Rectal cancer (CCR C20) T-stage=3 [Lower rectal cancer, ≤10cm from anal verge] OR [location of rectal cancer=low (SSW)] Low anterior resection (APDC: 32025-00, 32026-00, 32028-00))	No Yes: ClinCR Yes: DE No	83
Proportion of patients with 12 or more lymph nodes examined	Yes	"Number nodes examined" variable	Yes: DE	1,399 (67.0%)	Patients with DE	Yes: DE	2,088
Proportion of patients enrolled in a clinical trial	Yes	Date of first clinical trial enrollment	Yes	102 (2.4%)	All patients with ClinCR records that link to CCR-APDC records	Yes: ClinCR	4,244
Proportion of patients within 31-day target for time from decision to treat to commencement of treatment	Yes	Difference between *date of decision to treat* and first chemotherapy or radiotherapy treatment [Note: substantial number of negative differences]	Yes	443 (10.4%)	All patients with ClinCR records that link to CCR-APDC records	Yes: ClinCR	4,244
Proportion of patients within 62-day target for time from initial referral to commencement of treatment	Yes	Difference between *date of referral to cancer specialist* and first chemotherapy or radiotherapy treatment [Note: substantial number of negative differences]	Yes	1,438 (33.9%)	All patients with ClinCR records that link to CCR-APDC records	Yes: ClinCR	4,244
Proportion of major colorectal resections undertaken between 8pm and 8am	No	Time of procedure is not recorded on any data			All patients receiving surgery	No	28,360

DISCUSSION

The ClinCR dataset provides a range of data on cancer staging, clinical treatment and care processes that adds valuable information to other routinely collected data sources. The additional measures collected in the dataset extension provide more detailed information about colorectal-specific care and outcomes. While there are limitations in the coverage and quality of the ClinCR and extension data, without these datasets the development of population-based surgical indicators from routinely collected sources would be greatly limited.

Fifty-six percent of colorectal cancer cases registered in the CCR in 2007 and 2008 have a corresponding ClinCR record. The scope of collection of the current ClinCR means that cases of cancer in NSW residents who are diagnosed or treated solely in private facilities in NSW, or in public facilities in Greater Southern (GS) and Greater Western (GW) Area Health Services, or who are treated outside of NSW, are not included in the dataset. However, coverage for those with a procedure of interest in a public hospital in 2008, when all AHS had a well-established ClinCR, is over 94% for all AHS (see Table 5). There is a trend for increased coverage of all patients by year (see Table 3-4), which is also evident in the dramatic increase in coverage from the first year of operation to the subsequent year for the last AHS to establish a ClinCR. Even with the limitations of collection, those in the ClinCR appear to be largely representative of all patients treated for colorectal cancer in the time period (see Table 2). Nonetheless, the calculation of accurate rates of treatment and adherence to evidence-based guidelines is restricted by coverage of public facilities only, and by the exclusion of the two AHS (GS and GW).

Completeness is important for determining which items are reliable and provide useful information. Demographic and diagnostic items in the ClinCR have a high rate of completeness overall, with most non-optional items having data for over 95% of people (see Table 8-9). The four non-optional staging items provide complete data for 77% to 84% of people (see Table 10). Completeness of treatment items in the ClinCR is also mostly high. While treatment modality was specified in only 49% of records, all admission-related, radiotherapy and chemotherapy dates have data for 91% or more records except date of chemotherapy end (see Table 11-13). Specific items relating to treatment type or delivery range in completeness from 86% to 99%.

There is less information recorded for items relating to care, such as dates of referrals and consultations (see Table 14). For all of these items except psycho-social referral, less than 60% of people have data. However, it is difficult to determine whether the occasions of care did not take place or the data are simply missing. While many items in the ClinCR have a default option to code when the value of an item is unknown, it is hard to know whether these are used reliably. Similarly, some AHS have a higher number of records “flagged” as relating to a treatment event, but with lower completeness of individual treatment items. This variation may indicate that some AHS have different rules around the creation of treatment event records when specific data values are unknown. It is also an important consideration in the interpretation of rates of treatment and adherence to evidence-based guidelines.

Fifteen items in the ClinCR are duplicated or are closely related to items in other sources that formed our linked dataset (not including specific name and address identifiers). Mismatching values make up 3%

or less of data for all duplicate demographic items except for one country of birth item (ClinCR to APDC) (see Table 16). The lower rates of concordance between staging items, diagnostic items and admission-related dates are not ideal, but the size of the date differences are generally not large. Further investigation is needed to determine whether differences are inaccuracies or due to other reasons such as subsequent colorectal cancer diagnoses.

Date items in the ClinCR are mainly internally consistent. Overall, 82% of records do not contain any sets of illogical dates, and most of these records have less than two illogical date flags. Illogical date flags more often arise between dates of treatment or care and date of diagnosis, although there are greater median discrepancies between dates of treatment or care and date of death (see Table 17). While some discrepancies may be the result of incorrect recording, others will be accurate despite their unlikeliness.

Using routinely collected data to improve the timeliness and reduce the resource-intensiveness of clinical audit against best practice guidelines is only feasible if the data are able to provide enough information from which to develop indicators of outcome or evidence-based care. Of 34 proposed surgical indicators that were developed by an international collaboration, 21 could feasibly be calculated from items in the linked dataset (62%). ClinCR or extension items are required for 12 of these 21 indicators. However, six of these indicators rely on data from the colorectal dataset extension, which currently is only collected in two AHS. Within these AHS, completeness of dataset extension items range from 53% to 81% for core items (see Table 15).

Although the timeliness of construction of a linked dataset to enable monitoring of the uptake of evidence-based guidelines is limited by factors such as the availability of current cancer registration data and the time taken to receive and harmonise data from different datasets, there appears to be potential to conduct patterns of care studies using data routinely collected in NSW. Integration of the ClinCR into the CCR from July 2012 will facilitate this work.

Recommendations

The following recommendations arise solely from the analysis of data available for this report:

- To increase the coverage of the ClinCR to private facilities to enable better estimates of rates of treatment.
- To concentrate resources on data items that are unique to the ClinCR and not readily available from other existing data sources.
- To determine the accuracy of recording of items where doubt exists over whether data are missing or care or treatment did not occur e.g. through clinical record audit of a subset of the population.
- To standardise the recording of variables in each individual AHS ClinCR for ease of use.
- To increase the collection of the ClinCR and dataset extension to all AHS where they are not currently in operation.
- Research projects that require ClinCR data should take into consideration the coverage and completeness of records and should be restricted to patients diagnosed after 2007.
- Further assessment of the quality of ClinCR data for other tumour types for research purposes is recommended.

References

1. Tracey E, Kerr T, Dobrovic A, Currow D. Cancer in New South Wales: Incidence and mortality report 2008. Sydney: Cancer Institute NSW, August 2010

2. Center MM, Jemal A, Smith RA, Ward E. Worldwide variations in colorectal cancer. CA Cancer J Clin 2009, 59:366-378

3. Malin JL, Schneider EC, Epstein AM, Adams J, Emanuel EJ, Kahn KL. Results of the national initiative for cancer care quality: How can we improve the quality of cancer care in the United States? J Clin Oncol 2006, 24:626-634

4. Oliveria SA, Yood MU, Campbell UB, Yood SM, Stang P. Treatment and referral patterns for colorectal cancer. Med Care 2004, 42:901-906.

5. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA. The quality of health care delivered to adults in the United States. N Engl J Med 2003, 348:2635-2645.

6. Young JM, Leong D, Armstrong K, O'Connell D, Armstrong B, Spigelman A, Ackland S, Chapuis P, Kneebone A, Solomon MJ. Concordance with national guidelines for colorectal cancer care – a population-based patterns of care study of 3,096 patients. MJA 2007, 186:292-295

7. Cancer Institute NSW. NSW Colorectal Data Dictionary Version 1. Sydney, Australia, 2007

8. Roos LL, Mustard CA, Nicol JP, McLerran DF, Malenka DJ, Young TK, Cohen MM. Registries and administrative data: organisation and accuracy. Med Care 1993, 31:201-212

9. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chron Dis 1987, 40:373-383

10. Centre for Health Record Linkage (CHeRel). Information for researchers. Available at: www.cherel.org.au/researchers.html (last accessed 10 August 2010)

11. Hall SE, Holman CD, Platell C, Sheiner H, Threlfall T, Semmens J. Colorectal cancer surgical care and survival: do private health insurance, socioeconomic and locational status make a difference? ANZ J Surg 2005, 75:929-935

Appendix 1: Surgical procedures used to identify index admissions

ICD-10-AM code	Description
32000-00	Limited excision of large intestine with formation of stoma
32000-01	Right hemicolectomy with formation of stoma
32003-00	Limited excision of large intestine with anastomosis
32003-01	Right hemicolectomy with anastomosis
32004-00	Sub-total colectomy with formation of stoma
32004-01	Extended right hemicolectomy with formation of stoma
32005-00	Subtotal colectomy with ileosigmoid anastomosis
32005-01	Extended right hemicolectomy with anastomosis
32006-00	Left hemicolectomy with anastomosis
32006-01	Left hemicolectomy with formation of stoma
32009-00	Total colectomy with ileostomy
32012-00	Total colectomy with ileorectal anastomosis
32015-00	Total proctocolectomy with ileostomy
32024-00	High restorative anterior resection of rectum with intraperitoneal anastomosis
32025-00	Low restorative anterior resection of rectum with extraperitoneal anastomosis
32026-00	Low restorative anterior resection of rectum with coloanal anastomosis
32028-00	Ultra low restorative anterior resection of rectum with sutured coloanal anastomosis
32030-00	Rectosigmoidectomy with formation of stoma
32039-00	Abdominoperineal proctectomy
32047-00	Perineal proctectomy
32051-00	Total proctocolectomy with ileo-anal anastomosis
32051-01	Total proctocolectomy with ileo-anal anastomosis and formation of temporary ileostomy
32060-00	Closure of ileostomy with restoration of bowel continuity, with resection
32112-00	Perineal rectosigmoidectomy

ICD-10-AM code	Description
43993-01	Resection of large intestine or rectum for Hirschsprung's disease, with definitive anastomosis
32051-03	Total proctocolectomy with ileorectal anastomosis and formation of temporary ileostomy
43804-00	Resection of intestine for malrotation
43804-01	Resection of intestine for malrotation, with formation of stoma
43813-00	Repair of complicated meconium ileus with anastomosis
43813-01	Repair of complicated meconium ileus with stoma formation
43816-00	Resection of intestine for meconium ileus with anastomosis
43816-01	Resection of intestine for meconium ileus with stoma formation
43828-00	Resection of intestine for acute neonatal necrotising enterocolitis with anastomosis
43828-01	Resection of intestine for acute neonatal necrotising enterocolitis, with formation of stoma
43834-00	Resection of stricture of large intestine with anastomosis
43834-01	Resection of stricture of large intestine with formation of stoma
43936-00	Resection of intussusception, paediatric
43990-00	Resection of large intestine or rectum for Hirschsprung's disease, with stoma formation
90336-01	Isolation of segment of large intestine for interposition
Supplementary procedures used in Hall et al, ANZ J Surg, 2005[11]	
32099-00	Per anal submucosal excision of rectal tumour
32105-00	Per anal full thickness excision of anorectal tumour
32108-00	Trans-sphincteric excision of rectal tumour
43816-02	Repair of colonic atresia
43993-01	Resection of large intestine or rectum for Hirschsprung's disease, with definitive anastomosis
43993-02	Resection of large intestine or rectum for Hirschsprung's disease, with definitive anastomosis and stoma formation or revision
43996-00	Total colectomy for Hirschsprung's disease, with definitive anastomosis
90341-00	Excision of other rectal lesion

Appendix 2: Data Dictionary for collated ClinCR

Item	Values
All date items	DD/MM/YYYY Default date 01/01/1850 if date is unknown
Sex	1 - Male 2 - Female 9 - Not stated
Year of birth	NNNN
Country of birth	NNNN Standard Australian Classifications of Countries (SACC)
Postcode	NNNN 9990 - Overseas
Indigenous status	1 - Aboriginal 2 - Torres Strait Islander 3 - Both Aboriginal and Torres Strait Islander 4 - Neither Aboriginal or Torres Strait Islander 8 - Declined to respond 9 - Not stated/inadequately described
AMO registration number	Alphanumeric
Facility code	ANNN
Date of diagnosis of cancer	DD/MM/YYYY
Primary site of cancer	ICD-10-AM codes
Best basis for primary cancer diagnosis	1 - Histopathology 2 - Cytology 3 - Other 9 - Unknown
Histopathological grade	1 - Grade 1, Well differentiated, differentiated, NOS 2 - Grade 2, Moderately differentiated, moderately well differentiated, intermediate differentiation 3 - Grade 3, Poorly differentiated 4 - Grade 4, Undifferentiated, anaplastic 5 - T-cell (Lymphomas and leukaemias) 6 - B-cell, Pre-B, B-Precursor (Lymphomas and leukaemias) 7 - Null cell, Non T- non B (Lymphomas and leukaemias) 8 - Natural killer cell (Lymphomas and leukaemias) 9 - Grade/Differentiation unknown

Item	Values
Morphology of cancer	ICD-O-3 codes
Degree of spread of cancer	1 - Localised to the tissue of origin 2 - Invasion of adjacent tissue or organs 3 - Regional lymph nodes 4 - Distant metastases 5 - Not applicable 9 - Not known
Cancer staging - T stage code	X - Primary tumour cannot be assessed or is unknown 0 - No evidence of primary tumour IS - Carcinoma in situ 88 - TNM stage not applicable Staging 1 through 4D
Cancer staging - N stage code	X - Regional lymph nodes cannot be assessed or are unknown 0 - Nodes assessed, no evidence of regional lymph node metastasis 88 - TNM stage not applicable Staging 1 through 3C
Cancer staging - M stage code	X - Presence of distant metastases cannot be assessed or is unknown 0 - No known distant metastasis 88 - TNM stage not applicable Staging 1 though 1C
Staging basis	1 - Pathological 2 - Clinical 9 - Unknown
Other staging systems and classifications	(Relevant to colorectal cancer) 1 - TNM Classification of Malignant Tumours 4 - Australian Clinico-Pathological Staging (ACPS) System 5 - Duke's for colorectal cancers 23 - Modified Astler-Coller classification for colorectal 99 - Not for primary collection or unknown
Other stage groupings	Staging per above systems/classifications
Treatment modality	1 - Surgery 2 - Radiotherapy 3 - Medical Oncology and Haematology 4 - Diagnostic 5 - Admitted/Other
Treatment procedure for cancer (1 through 27)	ICD-10-AM codes

Item	Values
Radiotherapy treatment type	0 - No radiotherapy treatment given 1 - External beam radiation 2 - Brachytherapy 3 - Unsealed radioisotopes 9 - Radiotherapy administered but method not stated
Received radiation dose	NN in Grays (Gy) 0 - No radiation therapy was administered 9999 - Radiation therapy was administered but dose unknown
Radiation fractions	NN
Systemic therapy protocol name	eviQ ID where possible, CI-SCaT ID where no equivalent 0000 - No treatment 9999 - Unknown
Number of cycles of systemic therapy	NN 0 - Did not access planned cycles 98 - Not applicable 99 - Unknown
Performance status at diagnosis	0 - ECOG 0, Fully active, able to carry on all pre-disease performance without restriction 1 - ECOG 1, Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature e.g. light house work 2 - ECOG 2, Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about >50% of waking hours 3 - ECOG 3, Capable of only limited selfcare, confined to bed or chair >50% of waking hours 4 - ECOG 4, Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair 9 - Unknown
Psycho-social referral to	0 - Not indicated 1 - Psychiatrist 2 - Psychologist 3 - Social worker 4 - Specialist nurse / nurse counsellor 5 - Cancer support group or volunteer support group 7 - Counsellor / bereavement counsellor 8 - Pastoral care / chaplain / clergy / spiritual advisor 9 - Community service
Cause of death	ICD-10-AM site codes 6 - Non-cancer death 7 - Died, cause unknown

Appendix 3: Data Dictionary for collated extension items

Item	Values
Presentation	0 - Asymptomatic screening (FOBT) 1 – Symptoms of bowel cancer 2 - Emergency due to obstruction, perforation (with abscess/peritonitis), excessive bleeding or in conjunction with resuscitation 9 - Unknown
Method of surgery	0 - Open 1 - Laparoscopic / laparoscopic assisted 2 - Laparoscopic converted to open 3 - Local excision e.g. Transanal endoscopic microsurgery (TEMS) 8 - Other method of surgery 9 - No surgery
Level of rectal cancer	NN cm from anal verge
Residual tumour status	0 - No residual tumour (R0) 1 - Microscopic residual (R1) 2 - Macroscopic residual (R2) 9 - Residual status is unknown (RX)
Radial resection margin	NN
Lymphovascular invasion	0 - No 1 - Yes 9 - Unknown
Number of lymph nodes examined	NN 99 – Nodes removed but number of nodes unknown/not stated
Number of lymph nodes involved	NN 98 – No nodes examined 99 – Unknown if nodes are positive or negative
Mismatch repair deficiency (MMRD)	0 - No Staining (abnormal) 1 - Staining (normal) 2 - Not Done
Perineural invasion	0 - No invasion 1 - Invasion 8 - Invasion not assessed 9 - Unknown

Appendix 4: Completeness of ClinCR data by AHS

Demographic items

Sex

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS A	1,516 (99.9)	472 (99.8)	472 (99.8)
AHS B	1,628 (99.9)	576 (99.8)	576 (99.8)
AHS C	2,625 (100.0)	954 (100.0)	954 (100.0)
AHS D	2,367 (100.0)	1,232 (100.0)	1,232 (100.0)
AHS E	2,245 (100.0)	1,395 (100.0)	1,395 (100.0)
AHS F	2,585 (100.0)	821 (100.0)	821 (100.0)
Total	12,966 (100.0)	5,450 (100.0)	5,450 (100.0)

Year of birth

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS A	1,517 (100.0)	473 (100.0)	473 (100.0)
AHS B	1,628 (99.9)	576 (99.8)	576 (99.8)
AHS C	2,625 (100.0)	954 (100.0)	954 (100.0)
AHS D	2,367 (100.0)	1,232 (100.0)	1,232 (100.0)
AHS E	2,245 (100.0)	1,395 (100.0)	1,395 (100.0)
AHS F	2,585 (100.0)	821 (100.0)	821 (100.0)
Total	12,967 (100.0)	5,451 (100.0)	5,451 (100.0)

Country of birth

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS A	1,517 (100.0)	473 (100.0)	470 (99.4)
AHS B	1,618 (99.3)	573 (99.3)	572 (99.1)
AHS C	2,582 (98.4)	932 (97.7)	929 (97.4)
AHS D	2,349 (99.2)	1,221 (99.1)	1,219 (98.9)
AHS E	2,227 (99.2)	1,384 (99.2)	1,384 (99.2)
AHS F	2,583 (99.9)	819 (99.8)	818 (99.6)
Total	12,876 (99.3)	5,402 (99.1)	5,392 (98.9)

Postcode

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,517 (100.0)	473 (100.0)	466 (98.5)
AHS B	1,629 (100.0)	577 (100.0)	572 (99.1)
AHS C	2,624 (100.0)	953 (99.9)	934 (97.9)
AHS D	0 (0.0)	43 (3.5)	0 (0.0)
AHS E	2,245 (100.0)	1,395 (100.0)	1,357 (97.3)
AHS F	0 (0.0)	28 (3.4)	0 (0.0)
Total	8,015 (61.8)	3,469 (63.6)	3,329 (61.6)

Indigenous status

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,512 (99.7)	469 (99.2)	462 (97.7)
AHS B	1,622 (99.6)	574 (99.5)	570 (98.8)
AHS C	2,559 (97.5)	917 (96.1)	902 (94.5)
AHS D	2,346 (99.1)	1,218 (98.9)	1,195 (97.0)
AHS E	12 (0.5)	62 (4.4)	9 (0.6)
AHS F	2,579 (99.8)	818 (99.6)	794 (96.7)
Total	10,630 (82.0)	4,058 (74.4)	3,932 (72.1)

AMO registration number

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,511 (99.6)	472 (98.9)	469 (99.2)
AHS B	1,592 (97.7)	575 (99.7)	548 (95.0)
AHS C	2,561 (97.6)	909 (95.3)	894 (93.7)
AHS D	2,367 (100.0)	1,232 (100.0)	1,223 (99.3)
AHS E	2,117 (94.3)	1,290 (92.5)	1,282 (91.9)
AHS F	2,584 (100.0)	821 (100.0)	814 (99.1)
Total	12,732 (98.2)	5,299 (97.2)	5,230 (95.9)

Facility code

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,517 (100.0)	473 (100.0)	473 (100.0)
AHS B	1,627 (99.9)	577 (100.0)	574 (99.5)
AHS C	2,580 (98.3)	913 (95.7)	911 (95.5)
AHS D	2,367 (100.0)	1,232 (100.0)	1,231 (99.9)
AHS E	2,245 (100.0)	1,395 (100.0)	1,395 (100.0)
AHS F	2,585 (100.0)	821 (100.0)	821 (100.0)
Total	12,921 (99.6)	5,411 (99.2)	5,405 (99.1)

Diagnostic items

Date of diagnosis of cancer

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,517 (100.0)	473 (100.0)	473 (100.0)
AHS B	1,629 (100.0)	577 (100.0)	577 (100.0)
AHS C	2,625 (100.0)	954 (100.0)	954 (100.0)
AHS D	2,367 (100.0)	1,232 (100.0)	1,232 (100.0)
AHS E	2,245 (100.0)	1,395 (100.0)	1,395 (100.0)
AHS F	2,585 (100.0)	821 (100.0)	821 (100.0)
Total	12,968 (100.0)	5,452 (100.0)	5,452 (100.0)

Primary site of cancer

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,517 (100.0)	473 (100.0)	473 (100.0)
AHS B	1,629 (100.0)	577 (100.0)	577 (100.0)
AHS C	2,625 (100.0)	954 (100.0)	954 (100.0)
AHS D	2,367 (100.0)	1,232 (100.0)	1,232 (100.0)
AHS E	2,245 (100.0)	1,395 (100.0)	1,395 (100.0)
AHS F	2,585 (100.0)	821 (100.0)	821 (100.0)
Total	12,968 (100.0)	5,452 (100.0)	5,452 (100.0)

Best basis for primary cancer diagnosis

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,517 (100.0)	473 (100.0)	472 (99.8)
AHS B	1,628 (99.9)	577 (100.0)	576 (99.8)
AHS C	2,609 (99.4)	950 (99.6)	944 (99.0)
AHS D	2,367 (100.0)	1,232 (100.0)	1,230 (99.8)
AHS E	2,243 (99.9)	1,393 (99.9)	1,393 (99.9)
AHS F	2,585 (100.0)	821 (100.0)	821 (100.0)
Total	12,949 (99.9)	5,446 (99.9)	5,436 (99.7)

Histopathological grade

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,254 (82.7)	381 (80.5)	374 (79.1)
AHS B	1,386 (85.1)	488 (84.6)	481 (83.4)
AHS C	2,010 (76.6)	737 (77.3)	708 (74.2)
AHS D	2,069 (87.4)	1,072 (87.0)	1,050 (85.2)
AHS E	2,135 (95.1)	1,312 (94.1)	1,282 (91.9)
AHS F	2,030 (78.5)	664 (80.9)	648 (78.9)
Total	10,884 (83.9)	4,654 (85.4)	4,543 (83.3)

Morphology of cancer

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,517 (100.0)	473 (100.0)	473 (100.0)
AHS B	1,623 (99.6)	576 (99.8)	576 (99.8)
AHS C	2,624 (100.0)	953 (99.9)	953 (99.9)
AHS D	2,367 (100.0)	1,232 (100.0)	1,232 (100.0)
AHS E	2,245 (100.0)	1,395 (100.0)	1,395 (100.0)
AHS F	2,585 (100.0)	821 (100.0)	821 (100.0)
Total	12,961 (99.9)	5,450 (100.0)	5,450 (100.0)

Degree of spread of cancer

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,480 (97.6)	447 (94.5)	444 (93.9)
AHS B	1,573 (96.6)	554 (96.0)	549 (95.1)
AHS C	2,463 (93.8)	892 (93.5)	876 (91.8)
AHS D	2,327 (98.3)	1,213 (98.5)	1,200 (97.4)
AHS E	2,194 (97.7)	1,355 (97.1)	1,337 (95.8)
AHS F	2,387 (92.3)	753 (91.7)	740 (90.1)
Total	12,424 (95.8)	5,214 (95.6)	5,146 (94.4)

Staging items

Cancer staging - T stage code

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,281 (84.4)	377 (79.9)	367 (77.6)
AHS B	1,318 (80.9)	458 (79.4)	451 (78.2)
AHS C	1,847 (70.4)	679 (71.2)	655 (68.7)
AHS D	1,937 (81.8)	1016 (82.5)	993 (80.6)
AHS E	1,981 (88.2)	1212 (86.9)	1,184 (84.9)
AHS F	1,951 (75.5)	610 (74.3)	588 (71.6)
Total	10,315 (79.5)	4352 (79.8)	4,238 (77.7)

Cancer staging - N stage code

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,229 (81.0)	366 (77.4)	360 (76.1)
AHS B	1,312 (80.5)	456 (79.0)	451 (78.2)
AHS C	1,896 (72.2)	684 (71.7)	661 (69.3)
AHS D	1,919 (81.1)	1,010 (82.0)	989 (80.3)
AHS E	1,977 (88.1)	1,191 (85.4)	1,161 (83.2)
AHS F	1,865 (72.1)	591 (72.0)	572 (69.7)
Total	10,198 (78.6)	4,298 (78.8)	4,194 (76.9)

Cancer staging - M stage code

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,097 (72.3)	309 (65.3)	301 (63.6)
AHS B	1,365 (83.8)	455 (78.9)	448 (77.6)
AHS C	1,540 (58.7)	537 (56.3)	513 (53.8)
AHS D	2,060 (87.0)	1,042 (84.6)	1,019 (82.7)
AHS E	2,196 (97.8)	1,357 (97.3)	1,328 (95.2)
AHS F	1,782 (68.9)	513 (62.5)	494 (60.2)
Total	10,040 (77.4)	4,213 (77.3)	4,103 (75.3)

Staging basis

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,460 (96.2)	442 (93.4)	433 (91.5)
AHS B	279 (17.1)	127 (22.0)	111 (19.2)
AHS C	1,947 (74.2)	691 (72.4)	659 (69.1)
AHS D	2,274 (96.1)	1,176 (95.5)	1,156 (93.8)
AHS E	2,214 (98.6)	1,371 (98.3)	1,349 (96.7)
AHS F	2,504 (96.9)	790 (96.2)	787 (95.9)
Total	10,678 (82.3)	4,597 (84.3)	4,495 (82.4)

*Other staging systems and classifications**

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	974 (64.2)	327 (69.1)	317 (67.0)
AHS B	190 (11.7)	78 (13.5)	71 (12.3)
AHS C	17 (0.6)	29 (3.0)	5 (0.5)
AHS D	1,102 (46.6)	609 (49.4)	593 (48.1)
AHS E	2,113 (94.1)	1,308 (93.8)	1,260 (90.3)
AHS F	0 (0.0)	23 (2.8)	0 (0.0)
Total	4,396 (33.9)	2,374 (43.5)	2,246 (41.2)

* Optional item

*Other stage groupings**

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	929 (61.2)	295 (62.4)	284 (60.0)
AHS B	193 (11.8)	78 (13.5)	71 (12.3)
AHS C	17 (0.6)	29 (3.0)	5 (0.5)
AHS D	1,102 (46.6)	609 (49.4)	593 (48.1)
AHS E	2,107 (93.9)	1,302 (93.3)	1,254 (89.9)
AHS F	0 (0.0)	22 (2.7)	0 (0.0)
Total	4,348 (33.5)	2,335 (42.8)	2,207 (40.5)

* Dependent on above item

*TNM stage**

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	552 (36.4)	145 (30.7)	135 (28.5)
AHS B	0 (0.0)	12 (2.1)	0 (0.0)
AHS C	2,286 (87.1)	812 (85.1)	779 (81.7)
AHS D	2,071 (87.5)	1,045 (84.8)	1,019 (82.7)
AHS E	2,182 (97.2)	1,345 (96.4)	1,307 (93.7)
AHS F	0 (0.0)	25 (3.0)	0 (0.0)
Total	7,091 (54.7)	3,384 (62.1)	3,240 (59.4)

* Not in ClinCR data dictionary and not requested by researchers

*TNM edition number**

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,507 (99.3)	471 (99.6)	457 (96.6)
AHS B	0 (0.0)	8 (1.4)	0 (0.0)
AHS C	2,338 (89.1)	828 (86.8)	789 (82.7)
AHS D	0 (0.0)	15 (1.2)	0 (0.0)
AHS E	0 (0.0)	7 (0.5)	0 (0.0)
AHS F	0 (0.0)	3 (0.4)	0 (0.0)
Total	3,845 (29.6)	1,332 (24.4)	1,246 (22.9)

* Not in ClinCR data dictionary and not requested by researchers

Treatment items

Treatment modality

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS A	0 (0.0)	4 (0.8)	0 (0.0)
AHS B	1,627 (99.9)	577 (100.0)	577 (100.0)
AHS C	0 (0.0)	32 (3.4)	0 (0.0)
AHS D	0 (0.0)	29 (2.4)	0 (0.0)
AHS E	2,135 (95.1)	1,297 (93.0)	1,297 (93.0)
AHS F	2,585 (100.0)	821 (100.0)	821 (100.0)
Total	6,347 (48.9)	2,760 (50.6)	2,695 (49.4)

Admissions

	n records (% of records in ahs)	**n persons (% of persons in ahs)**
AHS A	1,135 (74.8)	436 (92.2)
AHS B	1,278 (78.5)	545 (94.5)
AHS C	1,949 (74.2)	849 (89.0)
AHS D	1,443 (61.0)	1,041 (84.5)
AHS E	1,250 (55.7)	1,203 (86.2)
AHS F	1,739 (67.3)	746 (90.9)
Total admissions	8,794 (67.8)	4,820 (88.4)

Date of admission

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS A	1,129 (99.5)	436 (100.0)	428 (98.2)
AHS B	1,270 (99.4)	545 (100.0)	538 (98.7)
AHS C	1,917 (98.4)	848 (99.9)	825 (97.2)
AHS D	1,412 (97.9)	1,034 (99.3)	1,006 (96.6)
AHS E	1,190 (95.2)	1,152 (95.8)	1,150 (95.6)
AHS F	1,730 (99.5)	746 (100.0)	736 (98.7)
Total	8,648 (98.3)	4,761 (98.8)	4,683 (97.2)

Date of discharge

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,129 (99.5)	436 (100.0)	428 (98.2)
AHS B	1,270 (99.4)	545 (100.0)	538 (98.7)
AHS C	1,879 (96.4)	835 (98.4)	793 (93.4)
AHS D	1,412 (97.9)	1,034 (99.3)	1,006 (96.6)
AHS E	1,190 (95.2)	1,152 (95.8)	1,150 (95.6)
AHS F	1,730 (99.5)	746 (100.0)	736 (98.7)
Total	8,610 (97.9)	4,748 (98.5)	4,651 (96.5)

*Procedure (ICD) version**

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,003 (88.4)	422 (96.8)	334 (76.6)
AHS B	0 (0.0)	4 (0.7)	0 (0.0)
AHS C	1,836 (94.2)	843 (99.3)	753 (88.7)
AHS D	0 (0.0)	19 (1.8)	0 (0.0)
AHS E	0 (0.0)	25 (2.1)	0 (0.0)
AHS F	1,583 (91.0)	736 (98.7)	618 (82.8)
Total	4,422 (50.3)	2,049 (42.5)	1,705 (35.4)

* Not in ClinCR data dictionary and not requested by researchers

*Treatment procedure for cancer (1)**

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	1,003 (88.4)	423 (97.0)	337 (77.3)
AHS B	1,117 (87.4)	541 (99.3)	438 (80.4)
AHS C	1,836 (94.2)	843 (99.3)	760 (89.5)
AHS D	1,443 (100.0)	1,041 (100.0)	1,038 (99.7)
AHS E	1,250 (100.0)	1,203 (100.0)	1,196 (99.4)
AHS F	1,583 (91.0)	736 (98.7)	638 (85.5)
Total	8,232 (93.6)	4,787 (99.3)	4,407 (91.4)

* There are 27 procedure items; this item reflects how many people have at least one procedure code recorded

Radiotherapy

	n records (% of records in ahs)	n persons (% of persons in ahs)
AHS A	72 (4.7)	66 (14.0)
AHS B	44 (2.7)	40 (6.9)
AHS C	111 (4.2)	100 (10.5)
AHS D	224 (9.5)	220 (17.9)
AHS E	166 (7.4)	160 (11.5)
AHS F	195 (7.5)	176 (21.4)
Total radiotherapy	812 (6.3)	762 (14.0)

Date of radiotherapy start

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	72 (100.0)	66 (100.0)	66 (100.0)
AHS B	44 (100.0)	40 (100.0)	40 (100.0)
AHS C	111 (100.0)	100 (100.0)	99 (99.0)
AHS D	224 (100.0)	220 (100.0)	220 (100.0)
AHS E	165 (99.4)	159 (99.4)	159 (99.4)
AHS F	144 (73.8)	134 (76.1)	127 (72.2)
Total	760 (93.6)	719 (94.4)	711 (93.3)

Date of radiotherapy end

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	68 (94.4)	64 (97.0)	62 (93.9)
AHS B	44 (100.0)	40 (100.0)	40 (100.0)
AHS C	103 (92.3)	93 (93.0)	92 (92.0)
AHS D	224 (100.0)	220 (100.0)	220 (100.0)
AHS E	165 (99.4)	159 (99.4)	159 (99.4)
AHS F	135 (69.3)	129 (73.3)	126 (71.6)
Total	739 (91.0)	705 (92.5)	699 (91.7)

Radiotherapy treatment type

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	71 (98.6)	66 (100.0)	65 (98.5)
AHS B	42 (95.5)	40 (100.0)	38 (95.0)
AHS C	80 (72.1)	78 (78.0)	76 (76.0)
AHS D	224 (100.0)	220 (100.0)	220 (100.0)
AHS E	166 (100.0)	160 (100.0)	160 (100.0)
AHS F	179 (91.8)	173 (98.3)	166 (94.3)
Total	762 (93.8)	737 (96.7)	725 (95.1)

Received radiation dose

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	67 (93.1)	64 (97.0)	61 (92.4)
AHS B	41 (93.2)	39 (97.5)	37 (92.5)
AHS C	78 (70.3)	76 (76.0)	73 (73.0)
AHS D	220 (98.2)	216 (98.2)	216 (98.2)
AHS E	166 (100.0)	160 (100.0)	160 (100.0)
AHS F	128 (65.5)	127 (72.2)	121 (68.8)
Total	700 (86.2)	682 (89.5)	668 (87.7)

Radiation fractions

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	68 (94.4)	64 (97.0)	62 (93.9)
AHS B	41 (93.2)	39 (97.5)	37 (92.5)
AHS C	78 (70.3)	76 (76.0)	73 (73.0)
AHS D	224 (100.0)	220 (100.0)	220 (100.0)
AHS E	166 (100.0)	160 (100.0)	160 (100.0)
AHS F	127 (65.1)	126 (71.6)	120 (68.2)
Total	704 (86.7)	685 (89.9)	672 (88.2)

Chemotherapy

	n records (% of records in ahs)	n persons (% of persons in ahs)
AHS A	316 (20.8)	229 (48.4)
AHS B	313 (19.2)	264 (45.8)
AHS C	552 (21.0)	300 (31.4)
AHS D	731 (30.9)	632 (51.3)
AHS E	719 (32.0)	542 (38.9)
AHS F	659 (25.5)	502 (61.1)
Total chemotherapy	3,290 (25.4)	2,469 (45.3)

Date of systemic therapy start

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	316 (100.0)	229 (100.0)	229 (100.0)
AHS B	312 (99.7)	263 (99.6)	263 (99.6)
AHS C	552 (100.0)	300 (100.0)	298 (99.3)
AHS D	731 (100.0)	632 (100.0)	627 (99.2)
AHS E	712 (99.0)	537 (99.1)	528 (97.4)
AHS F	479 (72.7)	370 (73.7)	344 (68.5)
Total	3,102 (94.3)	2,331 (94.4)	2,289 (92.7)

Date of systemic therapy end

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	269 (85.1)	202 (88.2)	185 (80.9)
AHS B	285 (91.1)	243 (92.0)	236 (89.4)
AHS C	462 (83.7)	265 (88.3)	229 (76.3)
AHS D	693 (94.8)	600 (94.9)	587 (92.9)
AHS E	707 (98.3)	535 (98.7)	523 (96.5)
AHS F	453 (68.7)	351 (69.9)	323 (64.3)
Total	2,869 (87.2)	2,196 (88.9)	2,083 (84.4)

Systemic therapy protocol name

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	315 (99.7)	228 (99.6)	228 (99.6)
AHS B	308 (98.4)	261 (98.9)	259 (98.1)
AHS C	550 (99.6)	298 (99.3)	298 (99.3)
AHS D	726 (99.3)	628 (99.4)	627 (99.2)
AHS E	719 (100.0)	542 (100.0)	542 (100.0)
AHS F	657 (99.7)	500 (99.6)	500 (99.6)
Total	3,275 (99.5)	2,457 (99.5)	2,454 (99.4)

Number of cycles of systemic therapy

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	248 (78.5)	175 (76.4)	165 (72.1)
AHS B	299 (95.5)	250 (94.7)	250 (94.7)
AHS C	479 (86.8)	268 (89.3)	241 (80.3)
AHS D	729 (99.7)	630 (99.7)	623 (98.6)
AHS E	693 (96.4)	529 (97.6)	511 (94.3)
AHS F	473 (71.8)	367 (73.1)	341 (67.9)
Total	2,921 (88.8)	2,219 (89.9)	2,131 (86.3)

Quality of care indicators

Date of referral to cancer specialist

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	195 (12.9)	135 (28.5)	17 (3.6)
AHS B	744 (45.7)	417 (72.3)	173 (30.0)
AHS C	302 (11.5)	197 (20.6)	36 (3.8)
AHS D	1,070 (45.2)	758 (61.5)	346 (28.1)
AHS E	437 (19.5)	296 (21.2)	41 (2.9)
AHS F	960 (37.1)	481 (58.6)	96 (11.7)
Total	3,708 (28.6)	2,284 (41.9)	709 (13.0)

Date of consultation with cancer specialist

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	279 (18.4)	182 (38.5)	33 (7.0)
AHS B	949 (58.3)	483 (83.7)	266 (46.1)
AHS C	469 (17.9)	270 (28.3)	68 (7.1)
AHS D	2,209 (93.3)	1,183 (96.0)	1,067 (86.6)
AHS E	847 (37.7)	553 (39.6)	98 (7.0)
AHS F	1,148 (44.4)	538 (65.5)	116 (14.1)
Total	5,901 (45.5)	3,209 (58.9)	1,648 (30.2)

Date of decision to treat

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	61 (4.0)	61 (12.9)	4 (0.8)
AHS B	41 (2.5)	43 (7.5)	1 (0.2)
AHS C	81 (3.1)	75 (7.9)	11 (1.2)
AHS D	220 (9.3)	211 (17.1)	35 (2.8)
AHS E	309 (13.8)	99 (7.1)	88 (6.3)
AHS F	132 (5.1)	129 (15.7)	4 (0.5)
Total	844 (6.5)	618 (11.3)	143 (2.6)

Date of first clinical trials enrolment

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS A	55 (3.6)	11 (2.3)	11 (2.3)
AHS B	184 (11.3)	54 (9.4)	53 (9.2)
AHS C	73 (2.8)	14 (1.5)	12 (1.3)
AHS D	47 (2.0)	22 (1.8)	17 (1.4)
AHS E	19 (0.8)	18 (1.3)	2 (0.1)
AHS F	73 (2.8)	13 (1.6)	12 (1.5)
Total	451 (3.5)	132 (2.4)	107 (2.0)

Date of first multidisciplinary team consultation (MDT)

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS A	154 (10.2)	32 (6.8)	30 (6.3)
AHS B	553 (33.9)	166 (28.8)	153 (26.5)
AHS C	790 (30.1)	228 (23.9)	215 (22.5)
AHS D	651 (27.5)	282 (22.9)	266 (21.6)
AHS E	419 (18.7)	232 (16.6)	212 (15.2)
AHS F	1,140 (44.1)	314 (38.2)	308 (37.5)
Total	3,707 (28.6)	1,254 (23.0)	1,184 (21.7)

Date of referral to palliative care

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS A	310 (20.4)	74 (15.6)	68 (14.4)
AHS B	304 (18.7)	70 (12.1)	66 (11.4)
AHS C	825 (31.4)	219 (23.0)	205 (21.5)
AHS D	878 (37.1)	378 (30.7)	363 (29.5)
AHS E	292 (13.0)	196 (14.1)	171 (12.3)
AHS F	1,048 (40.5)	220 (26.8)	209 (25.5)
Total	3,657 (28.2)	1,157 (21.2)	1,082 (19.8)

Performance status at diagnosis

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS A	444 (29.3)	95 (20.1)	86 (18.2)
AHS B	544 (33.4)	152 (26.3)	147 (25.5)
AHS C	450 (17.1)	158 (16.6)	136 (14.3)
AHS D	801 (33.8)	347 (28.2)	321 (26.1)
AHS E	952 (42.4)	554 (39.7)	527 (37.8)
AHS F	315 (12.2)	80 (9.7)	66 (8.0)
Total	3,506 (27.0)	1,386 (25.4)	1,283 (23.5)

Psycho-social referral to

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS A	713 (47.0)	188 (39.7)	179 (37.8)
AHS B	1,587 (97.4)	559 (96.9)	552 (95.7)
AHS C	1,644 (62.6)	507 (53.1)	474 (49.7)
AHS D	2,358 (99.6)	1,225 (99.4)	1,199 (97.3)
AHS E	973 (43.3)	597 (42.8)	566 (40.6)
AHS F	2,382 (92.1)	710 (86.5)	693 (84.4)
Total	9,657 (74.5)	3,786 (69.4)	3,663 (67.2)

Date of death

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS A	629 (41.5)	159 (33.6)	156 (33.0)
AHS B	605 (37.1)	178 (30.8)	174 (30.2)
AHS C	1,340 (51.0)	372 (39.0)	361 (37.8)
AHS D	0 (0.0)	24 (1.9)	0 (0.0)
AHS E	444 (19.8)	295 (21.1)	282 (20.2)
AHS F	1,156 (44.7)	277 (33.7)	267 (32.5)
Total	4,174 (32.2)	1,305 (23.9)	1,240 (22.7)

Cause of death

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS A	0 (0.0)	0 (0.0)	0 (0.0)
AHS B	214 (13.1)	81 (14.0)	76 (13.2)
AHS C	0 (0.0)	5 (0.5)	0 (0.0)
AHS D	0 (0.0)	4 (0.3)	0 (0.0)
AHS E	442 (19.7)	283 (20.3)	268 (19.2)
AHS F	0 (0.0)	3 (0.4)	0 (0.0)
Total	656 (5.1)	376 (6.9)	344 (6.3)

Appendix 5: Completeness of colorectal dataset extension by AHS

	n records (% of records in ahs)	n persons (% of persons in ahs)
AHS X	[redacted] (92.7)	[redacted] (92.4)
AHS Y	[redacted] (54.3)	[redacted] (52.1)
Total extension	3,413 (26.3)	1,908 (35.0)

* Blacked out to preserve AHS identity

Presentation

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS X	1,628 (74.2)	793 (68.5)	787 (68.0)
AHS Y	1,013 (83.1)	591 (78.7)	584 (77.8)
Total	2,641 (77.4)	1,384 (72.5)	1,371 (71.9)

Method of surgery

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS X	760 (34.6)	395 (34.1)	393 (34.0)
AHS Y	1,057 (86.7)	632 (84.2)	630 (83.9)
Total	1,817 (53.2)	1027 (53.8)	1,023 (53.6)

*Level of rectal cancer**

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS X	277 (12.6)	130 (11.2)	126 (10.9)
AHS Y	157 (12.9)	56 (7.5)	55 (7.3)
Total	434 (12.7)	186 (9.7)	181 (9.5)

* Number of records (persons) with rectal cancer diagnosis = 1,217 (581), rectosigmoid = 275 (151), total = 1,492 (732)

Residual tumour status

	Records with data, n (% in ahs)	Persons with *any* data, n (% in ahs)	Persons with *all* data, n (% in ahs)
AHS X	1,577 (71.9)	834 (72.1)	830 (71.7)
AHS Y	1,033 (84.7)	618 (82.3)	615 (81.9)
Total	2,610 (76.5)	1,452 (76.1)	1,445 (75.7)

Radial resection margin

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS X	387 (17.6)	176 (15.2)	172 (14.9)
AHS Y	794 (65.1)	488 (65.0)	484 (64.4)
Total	1,181 (34.6)	664 (34.8)	656 (34.4)

Lymphovascular invasion

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS X	1,659 (75.6)	885 (76.5)	883 (76.3)
AHS Y	978 (80.2)	589 (78.4)	586 (78.0)
Total	2,637 (77.3)	1,474 (77.3)	1,469 (77.0)

Number of lymph nodes examined

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS X	1,770 (80.7)	937 (81.0)	935 (80.8)
AHS Y	981 (80.5)	597 (79.5)	594 (79.1)
Total	2,751 (80.6)	1,534 (80.4)	1,529 (80.1)

Number of lymph nodes involved

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS X	1,773 (80.8)	939 (81.2)	935 (80.8)
AHS Y	987 (81.0)	600 (79.9)	597 (79.5)
Total	2,760 (80.9)	1,539 (80.7)	1,532 (80.3)

Mismatch repair deficiency (MMRD)

	Records with data, n (% in ahs)	**Persons with *any* data, n (% in ahs)**	**Persons with *all* data, n (% in ahs)**
AHS X	1,670 (76.1)	881 (76.1)	879 (76.0)
AHS Y	1,029 (84.4)	620 (82.6)	617 (82.2)
Total	2,699 (79.1)	1,501 (78.7)	1,496 (78.4)