



NATIONAL BOWEL CANCER AUDIT

Using cancer registry data to improve case ascertainment

NBOCA: Short report

Date of publication: 11 June 2020

About HQIP, the National Clinical Audit and Patient Outcomes Programme and how it is funded:

The Healthcare Quality Improvement Partnership (HQIP) is led by a consortium of the Academy of Medical Royal Colleges, the Royal College of Nursing and National Voices. Its aim is to promote quality improvement in patient outcomes, and in particular, to increase the impact that clinical audit, outcome review programmes and registries have on healthcare quality in England and Wales. HQIP holds the contract to commission, manage and develop the National Clinical Audit and Patient Outcomes Programme (NCAPOP), comprising around 40 projects covering care provided to people with a wide range of medical, surgical and mental health conditions. The programme is funded by NHS England, the Welsh Government and, with some individual projects, other devolved administrations and crown dependencies. www.hqip.org.uk/national-programmes

Executive Summary

Initial linkage between cancer registration data provided by the National Cancer Registration and Analysis Service (NCRAS) and clinical audit (NBOCA) data for patients in England published in the 2019 Annual Report indicated there was a difference in the number and type of patients captured by the data sources. This report aims to understand the reasons for this difference in captured cases and to determine whether NCRAS data can be routinely linked to NBOCA data to improve case ascertainment.

Between April 2016 and March 2017 4,200 fewer patients were captured by NBOCA than NCRAS. The reasons behind this are multifactorial. Ninety percent of NBOCA patients diagnosed in this time period and almost eighty percent of NCRAS patients could be linked via NHS number.

Unlinked NCRAS patients were more likely to have advanced cancer, to present in an emergency, and to have much higher mortality rates with over a third dying within 3 months of diagnosis. Many of these patients are likely to have had little contact with secondary care and are therefore less likely to have been entered into NBOCA.

A small number of NHS Trusts submitted $\leq 60\%$ of expected NCRAS patients to NBOCA for the time period April 2017 to March 2018. The majority of missing cases are distributed across NHS Trusts.

Routine linkage to NCRAS would allow NBOCA to report with complete case capture, but requires improvement to linkage methods, assessment of the impact of decreased data completeness and acknowledgement that cohorts compared are not exact due to differences in recorded date of diagnosis.

Introduction

The National Bowel Cancer Audit (NBOCA) was established in its current form in 2005. It was established as an audit of surgical practice in colorectal (bowel) cancer patients. Since 2015, the audit has expanded its scope and now also includes other treatments provided in secondary (hospital) care. It aims to assess the process of care and its outcomes in all adults diagnosed with bowel cancer in England and Wales. Data is routinely collected via patient cancer management systems at MDT meetings and transferred to a database (Clinical Audit Platform CAP) held by the Clinical Audit and Registries Management Service (CARMS) at NHS Digital.

The National Cancer Registration and Analysis Service (NCRAS) is run by Public Health England (PHE) and collects standardised information on over 300,000 cases of cancer diagnosed each year in England via multiple sources including healthcare records e.g. hospital/GP and death certificates.⁽¹⁾

Audit and NCRAS data were first linked together for the NBOCA 2019 Annual Report.⁽²⁾ This was to enable the audit to compare information about the diagnosis and treatment recorded for patients and to quantify the number of patients who were not being entered into the audit. Chapter 7 of this report compared patients with a recorded diagnosis date between April 2017 and March 2018 in each dataset and found substantially more patients in NCRAS than

in audit data. There were also differences in the demographic/tumour staging between NCRAS patients who did and did not link.

The initial linkage work performed for the 2019 Annual Report indicated there was often a difference in recorded date of diagnosis in patients who were present in both datasets. If the difference in dates meant that patients were allocated into different audit years of diagnosis (April to March) then restricting both datasets to patients diagnosed in one audit year could potentially affect the linkage rate. In addition, patients with type 2 objections/national data opt out (objection to personal confidential information being used for purposes other than their direct care registered with NHS Digital)⁽³⁾ were removed from the NCRAS data, again reducing how many audit records could be linked. If there were no data linkage errors then we would expect no patients in the audit who could not be linked to NCRAS, other than those with a type 2 objections/national data opt out, because NCRAS has complete capture of cases.

This report aims to understand the reasons for the differences in cases captured in the two datasets and to determine whether NCRAS data can be routinely linked to audit data to provide audit reporting with improved ascertainment. For this to be possible, linkage of the two datasets would need to have a low rate of missed links to avoid the same patients being reported on twice. Also, data completeness of key data items in NCRAS data would need to be good.

The objectives of the report are therefore to:

- (i) investigate how much allowing linkage to patients with a broader range of diagnosis dates improves the linkage rate
- (ii) assess the quality of the linkage of the two datasets by carrying out clerical linkage of the unlinked cases, and quantifying the extent of national data opt-out
- (iii) identify which patient groups are under-captured in NBOCA
- (iv) assess the data completeness of key data items in NCRAS records not linked to NBOCA
- (v) assess the distribution of missing patients across hospital trusts

Methods

(i) Methods to improve the linkage rate

NCRAS data for patients with an ICD-10 diagnosis code C18-C20 and a date of diagnosis between April 2010 and March 2018 was requested from PHE. Eligible audit patients were those with a date of diagnosis within the same time period; those without a type 2/national data opt out registered with NHS Digital had their details sent to PHE for linkage to the NCRAS dataset.

The two datasets were linked only using NHS number by PHE, and the NCRAS dataset containing linked and unlinked patients with colorectal cancer returned. Any remaining patients with type2/national data opt out were removed before the dataset left PHE.

In order to account for differences in diagnosis date at either end of the audit year, comparisons were made on patients diagnosed between April 2016 and March 2017 rather than the patients included in the 2019 Annual Report (diagnosed between April 2017 and March 2018). Shifting to an earlier reporting period meant that there were NCRAS data with diagnosis dates before and after the reporting period. Initial data restrictions were the same as those in the report, only patients aged over 18 years with a diagnosis recorded in England were eligible. Carcinoid/neuroendocrine tumours were excluded from NCRAS data to be consistent with the audit, and appendiceal tumours (C181) were removed from both datasets in an attempt to ensure all remaining carcinoid tumours were excluded.

In an attempt to improve the linkage rate estimated in the annual report, data for audit patients diagnosed between April 2016 and March 2017 were linked to patients diagnosed between April 2015 and March 2018 in NCRAS, and vice versa.

(ii) Methods to assess linkage quality

We used “clerical linkage” to attempt to link the remaining unlinked 2016/17 audit patients to unlinked 2015-2018 NCRAS patients using variables present in both datasets. The variables used to link datasets were the diagnosing trust, sex, Lower Layer Super Output Area (LSOA) of residence, ICD-10 code of tumour location, age, date of diagnosis, and date of death. LSOA's are small geographical areas of England and Wales consisting of a mean population of 1500.

Where only one NCRAS and audit record matched on diagnosing trust, sex, LSOA, age +/-1 year and date of diagnosis +/- 90 days, these were considered a link if tumour sites in the two datasets were not on opposite sides of the bowel, dates of death did not disagree, and dates of death were not before dates of diagnosis between data sources. Where there were multiple potential NCRAS records linking to an audit record, or multiple potential audit records linking to a NCRAS record, the record-pair with the greatest number of matching variables was chosen. This is a conservative linkage algorithm which minimises false links but is likely to miss links.

To quantify the effect of type 2 objections/national data opt out on linkage quality, a variable was created in the NBOCA dataset indicating whether the patient was present in the dataset when the data was sent for ONS/HES linkage, at which point type 2 objections/national data opt outs were upheld. The presence of this variable with no linkage to ONS was used as a proxy for the existence of a type 2 objection/national data opt out.

(iii) Methods to identify which patients are under-captured in NBOCA

Patients not entering secondary care are not expected to be captured in the audit. These patients were identified as those diagnosed at a trust that does not provide secondary care for bowel cancer, and patients in NCRAS without a diagnosing trust e.g. record of cancer obtained from a death certificate only. The number of these patients in NCRAS was quantified. In addition, the characteristics of NCRAS patients not linked to the audit was examined to identify further groups of patients under-captured in the audit.

(iv) Methods to assess the data completeness of key data items in patients in NCRAS not linked to NBOCA

The data completeness of the following key data items was assessed in NCRAS patients not linked to the audit: age, sex, index of multiple deprivation based on patient's postcode of residence (IMD), source of referral, location of tumour and tumour stage.

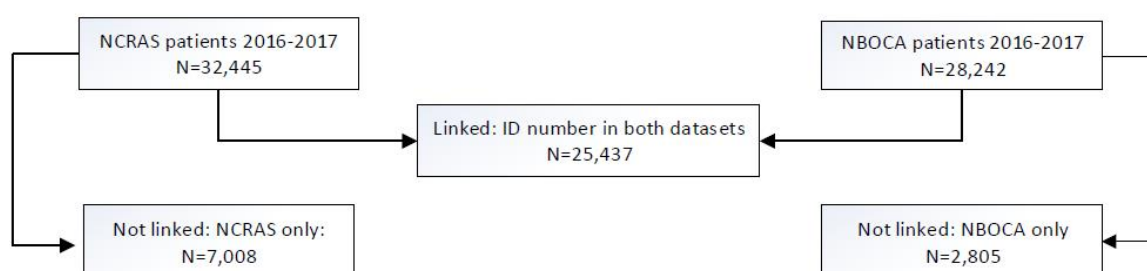
(v) Methods to assess the distribution of missing patients across hospital trusts

To quantify the distribution of missing cases across trusts, the same exclusions described above (under 18s, diagnosed outside England, carcinoid/neuroendocrine/appendiceal tumours) were applied to the 2017/18 datasets used in the 2019 Annual Report, and the number of trusts who had poor NCRAS case ascertainment obtained. Poor case ascertainment was defined as NBOCA patient total less than 60% of NCRAS patient total.

Results

The results of the initial linkage of patients with a recorded diagnosis date between April 2016 and March 2017 are shown in Figure 1. The audit captured approximately 4,200 fewer patients than NCRAS (87% case ascertainment in the audit compared to NCRAS). This compares to an overall case ascertainment of 92% using HES/PEDW as the denominator, as published in the 2018 Annual Report. Ninety percent of audit patients diagnosed in this time period and almost 80% of NCRAS patients could be linked.

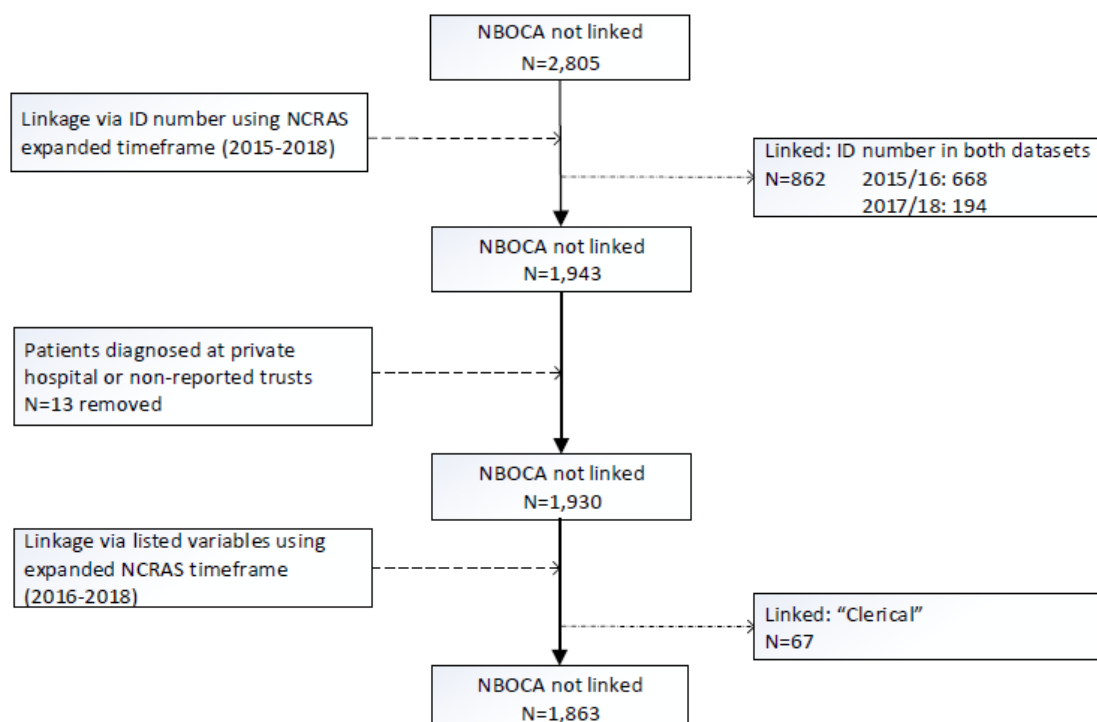
Figure 1: Flow Chart showing linkage of NCRAS and NBOCA patients diagnosed between April 2016 and March 2017 following initial exclusions



(i) Improving the linkage rate by allowing linkage to patients with a broader range of diagnosis dates

Just over 30% of the 2,805 unlinked 2016/17 audit patients could be linked via audit ID to an NCRAS record between 2015 and 2018, almost 80% of whom had a diagnosis date in the earlier audit year (Figure 2).

Figure 2: Flow Chart showing further linkage of NBOCA patients unlinked after initial linkage



Almost 20% of the 7,008 unlinked 2016/17 NCRAS patients could be linked to an audit record by broadening the range of diagnosis dates in the audit (Figure 3).

(ii) Assessing the quality of the linkage of the two datasets by carrying out clerical linkage of the unlinked cases, and quantifying the extent of national patient opt-out

As explained, the linkage algorithm was conservative to minimise the number of false links, and the proportion of unlinked records which are linked by this conservative clerical linkage method gives a lower bound for the estimate of the rate of missed links. A further small percentage of unlinked audit records (67/1,930 records) and unlinked NCRAS records (57/5,127 records) could be linked by clerical linkage (Figures 2 and 3), corresponding to minimum estimates of the missed linkage rate of 3.5% and 1.1% respectively.

Around a further one third of audit patients who did not initially link to NCRAS (912) did not have ONS data returned and could be assumed to have an opt-out in place. This group formed almost half (812) of the final 1,863 unlinked audit patients. Overall, the 912 patients form 3.2% of the initial audit 2016-2017 patient total (28,242). Audit patients with an assumed data-opt out were more likely to be aged 65-84 years, less deprived and more likely to be a GP referral (data not shown).

(iii) Identifying which patient groups are under-captured in NBOCA

7% of unlinked NCRAS records did not have a diagnosing trust code recorded and so would not be expected to be entered into the audit (Figure 3).

Figure 3: Flow Chart showing further linkage of NCRAS patients unlinked after initial linkage

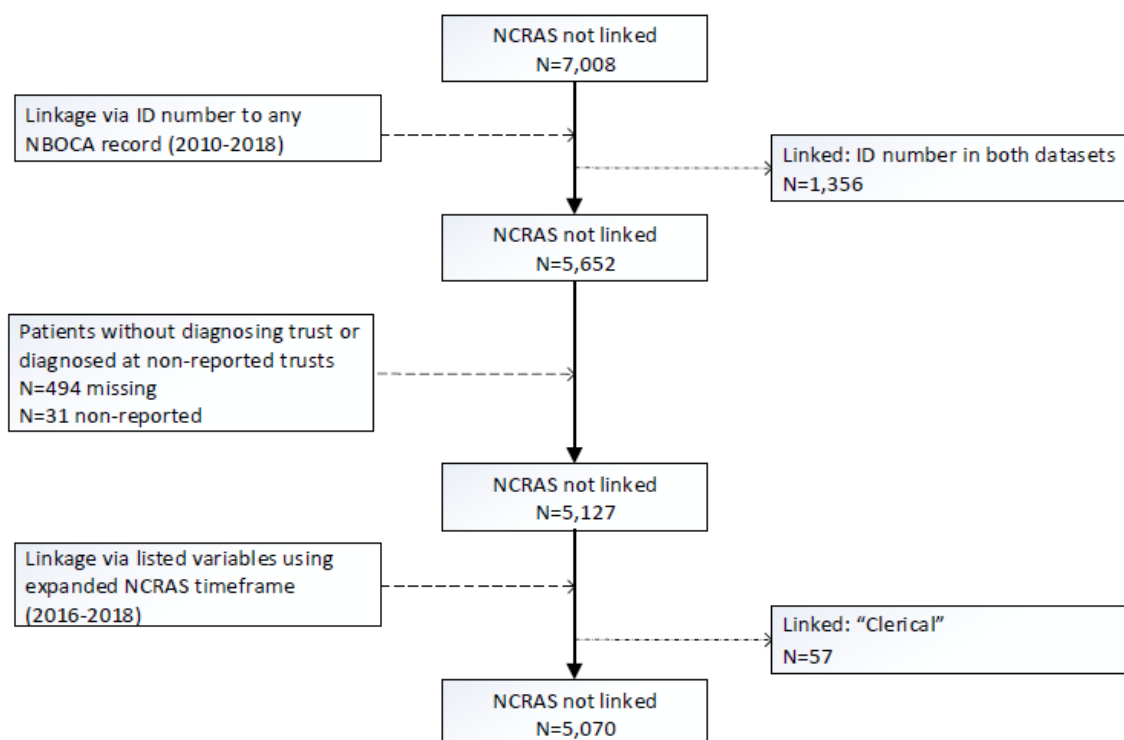


Table 1 shows patient demographic and tumour information for patients linked between the datasets, unlinked audit and unlinked NCRAS patients. The data for linked patients comes from the audit; the source of mortality information for audit patients is linked NBOCA-ONS data and that recorded in NCRAS for unlinked NCRAS patients.

Compared to linked records, unlinked audit records were more representative than unlinked NCRAS records. Over half of unlinked audit patients were GP referrals, with around 15% diagnosed after an emergency admission. As shown above, around half of unlinked audit patients can be assumed to have had a type-2/national data opt-out and these patients are likely to be reasonably representative of all patients.

The unlinked NCRAS patients were far more likely to be diagnosed after an emergency admission. Tumour stage was similar between the linked records and unlinked audit records, but substantially more advanced for the unlinked NCRAS records. Unlinked patients also tended to be older with over 40% of unlinked audit patients and over 50% of unlinked NCRAS patients aged 75 or older at time of diagnosis. A slightly higher proportion of female patients were unlinked which is likely to be related to their older age. Approximately 15% of the unlinked NCRAS records had their tumour location recorded as being in an unspecified location of the colon (C18.9) or in overlapping sites of the colon (C18.8), which are excluded from the audit.

The mortality in unlinked patients is substantially higher at all follow-up times, with much larger differences in the unlinked NCRAS records than unlinked audit records. This would fit with the higher proportions of cases with advanced disease and emergency admissions, i.e. disease that leads to death soon after diagnosis and them subsequently not being discussed at MDT (which should automatically lead to an NBOCA submission), and also with the more advanced age of these patients.

Table 1 Patient information for linked and unlinked NBOCA-NCRAS patients

		Linked NBOCA and NCRAS		Unlinked NBOCA		Unlinked NCRAS	
		N	%	N	%	N	%
Overall		26,366		1,876		5,595	
Age group (years)	<50	1,463	5.5	133	7.1	256	4.6
	50-64	5,909	22.4	386	20.6	887	15.9
	65-74	7,936	30.1	573	30.5	1,286	23.0
	75-84	7,774	29.5	553	29.5	1,627	29.1
	>=85	3,284	12.5	231	12.3	1,539	27.5
Gender	Male	15,071	57.2	1,041	55.5	2,914	52.1
	Female	11,280	42.8	834	44.5	2,681	47.9
	Missing	15	0.1	1	0.1	0	0.0
Quintile of Deprivation	5 most deprived	4,097	15.5	312	16.6	958	17.1
	4	4,787	18.2	319	17.0	1,065	19.0
	3	5,441	20.6	381	20.3	1,197	21.4
	2	5,982	22.7	438	23.3	1,192	21.3
	1 least deprived	6,026	22.9	407	21.7	1,183	21.1
	Missing	33	0.1	19	1.0	0	0.0
Source of referral	Emergency Admission	4,024	15.3	322	17.2	1,513	27.0
	GP Referral	14,621	55.5	982	52.3	971	17.4
	Screening Referral	2,674	10.1	129	6.9	224	4.0
	Other	4,569	17.3	402	21.4	1,082	19.3
	Not known	478	1.8	28	1.5	1,805	32.3
Tumour location	Ascending colon	2,854	10.8	180	9.6	504	9.0
	Caecum	3,909	14.8	346	18.4	784	14.0
	Colon, NOS	n/a	n/a	n/a	n/a	790	14.1
	Descending colon	961	3.6	67	3.6	158	2.8
	Hepatic flexure	1,080	4.1	58	3.1	187	3.3
	Overlapping lesion	n/a	n/a	n/a	n/a	26	0.5
	Rectosigmoid junction	1,481	5.6	96	5.1	362	6.5
	Rectum, NOS	7,395	28.0	594	31.7	1,289	23.0
	Sigmoid colon	6,191	23.5	379	20.2	1,059	18.9
	Splenic flexure	702	2.7	50	2.7	113	2.0
	Transverse colon	1,793	6.8	106	5.7	323	5.8
Tumour stage	1	4,052	15.4	259	13.8	671	12.0
	2	4,474	17.0	268	14.3	668	11.9
	3	7,664	29.1	423	22.5	846	15.1
	4	4,851	18.4	381	20.3	1,754	31.4
	Missing	5,325	20.2	545	29.1	1,656	29.6
Mortality from diagnosis	Mortality denominator	26,137*		955*		5,595	
	7 days	312	1.2	15	1.6	679	12.1
	30 days	957	3.7	57	6.0	1,286	23.0
	90 days	2,439	9.3	125	13.1	2,001	35.8
	182 days	3,745	14.3	205	21.5	2,527	45.2
	365 days	5,678	21.7	302	31.6	3,021	54.0

* Source of mortality information is linked NBOCA-ONS. ONS mortality information is not available for all patients due to assumed data opt-out

(iv) Assessing the data completeness of key data items in NCRAS records not linked to NBOCA

The unlinked NCRAS data had a substantially higher proportion of missing data for referral source and cancer stage (approximately 30% for both).

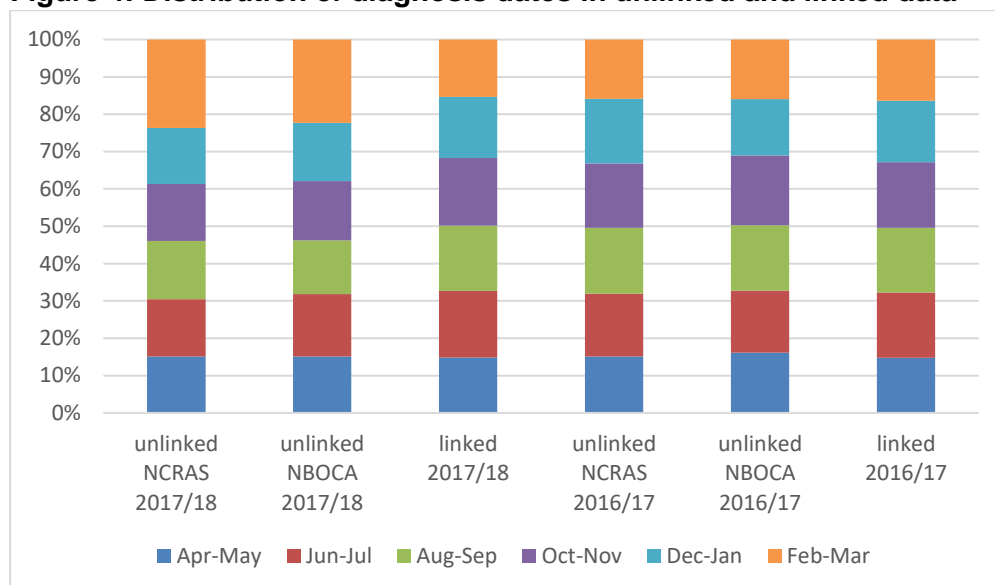
(v) Assessing the distribution of missing patients across hospital trusts

The initial case ascertainment comparing totals of patients submitted in the audit compared to NCRAS in the 2017/18 data used for the 2019 Annual Report, was 91.6%. Note that there was higher case ascertainment in 2017/18 data than in 2016/17 data. In 2017/18 data, 5 trusts were identified as having poor case ascertainment compared to NCRAS, defined as NBOCA patient total less than 60% of NCRAS patient total.

Following removal of patients who were not diagnosed at a Trust providing secondary care for bowel cancer (patients diagnosed at a trust that does not provide secondary care for bowel cancer or without a recorded diagnosing trust), case ascertainment compared to NCRAS was 92.9%. Excluding the 5 trusts with poor NBOCA submission data would increase the case ascertainment to 94.5%.

In both the final NCRAS and audit unlinked 2017/18 patients, a greater proportion of the overall total had a diagnosis date between January and March 2018: around 30% when 25% would be expected with an even distribution across the year. The difference is greatest in patients recorded in either dataset as being diagnosed in March 2018. The distribution of diagnosis dates between linked and unlinked data in 2017/18 and 2016/17 is shown in Figure 4.

Figure 4: Distribution of diagnosis dates in unlinked and linked data



Conclusions

The audit captures fewer cases of bowel cancer than NCRAS, with a 92% case ascertainment in the most recent audit year. It is important to understand why there are patients missing from the audit and which groups of patients are under-captured, to assess how these patients can be captured, and whether they should be included in an audit of secondary care. 7% of unlinked NCRAS records did not have a diagnosing trust code recorded and so would not be expected to be entered into the audit. 15% of NCRAS patients who were not linked to the audit were recorded as having cancer of an unspecified location of the colon or overlapping sites of the colon, which are excluded from the audit.

Unlinked NCRAS patients were more likely to have advanced cancer, to present in an emergency, and to have much higher mortality rates with over a third dying within 3 months of diagnosis. Many of these patients are likely to have had little contact with secondary care and are therefore less likely to have been entered into the audit. These findings are consistent with those recently published by the National Oesophago-Gastric Cancer Audit.⁽⁴⁾

Only 5 hospital trusts had very poor case ascertainment compared to NCRAS (less than 60%). The vast majority of cases missing from the audit are distributed across hospital trusts.

Routine linkage to NCRAS would allow the audit to report with complete case capture, but this would require minimal rates of linkage error to avoid reporting on the same patient twice (patient present in both datasets without linkage), and reasonable data completeness of key data items.

The findings of this report demonstrate that at least 1% of NCRAS patients who are not linked to the audit via NHS number are missed links, and this number is likely to be higher with less conservative linkage methods. Improved data linkage methods, such as probabilistic linkage using more patient identifiers, could be used to ensure the number of missed links is minimised. Type 2 objections /national data opt out affected the linkage rate between the audit and NCRAS but from 2020 these patients will be removed from both datasets, and will therefore affect case ascertainment rather than data linkage. Differences in the recorded date of diagnosis between datasets means that the group of patients compared will never be exactly the same, and can affect linkage rates if date restrictions are used to define the patient cohorts from the two datasets.

The data completeness of some key data items in NCRAS patients unlinked to the audit is poor. Cancer stage is a particularly important data item for the audit as it is required for describing care pathways and for risk adjustment. Linkage to NCRAS would increase case ascertainment whilst decreasing data completeness, and the impact of this will need to be assessed.

Implications for the audit

1. NCRAS data should be routinely linked to audit data as long as improved data linkage methods are used to minimise the linkage errors
2. The impact of routine linkage of audit data to NCRAS data on data completeness should be assessed and monitored
3. Patients with cancer of an unspecified location of the colon or in overlapping sites of the colon (C18.8 or C18.9) should be included in NBOCA
4. Trusts with poor case ascertainment should work to improve their submission of data to NBOCA
5. Further exploration of the care pathways of patients should be used to determine whether it is appropriate to include all patients captured in NCRAS but not in the audit in an audit of secondary care

References

1. National Cancer Registration and Analysis Service (NCRAS) Guidance. [20th February 2020]; Available from: <https://www.gov.uk/guidance/national-cancer-registration-and-analysis-service-ncras>.
2. National Bowel Cancer Audit 2019 Annual Report. [January 2020]; Available from: <https://www.nboca.org.uk/reports/annual-report-2019/>.
3. National data opt-out. [6th March 2020]; Available from: <https://digital.nhs.uk/services/national-data-opt-out>
4. National Oesophago-Gastric Cancer Audit Short Report Comparison of patients captured by NOGCA and the National Cancer Registration and Analysis Service. [June 2020]; Available from: <https://www.nogca.org.uk/reports/>.

This project involves data derived from patient-level information collected by the NHS, as part of the care and support of cancer patients. The data is collated, maintained and quality assured by the National Cancer Registration and Analysis Service, which is part of Public Health England (PHE). Access to the data was facilitated by the PHE Office for Data Release.