Open access Research

# BMJ Open Post-colonoscopy colorectal cancers identified by probabilistic and deterministic linkage: results in an Australian prospective cohort

Kavitha Subramaniam, <sup>1,2</sup> P W Ang, <sup>3</sup> Teresa Neeman, <sup>4</sup> Mitali Fadia, <sup>5</sup> Doug Taupin <sup>1,3</sup>

To cite: Subramaniam K. Ang PW. Neeman T. et al. Postcolonoscopy colorectal cancers identified by probabilistic and deterministic linkage: results in an Australian prospective cohort. BMJ Open 2019;9:e026138. doi:10.1136/ bmjopen-2018-026138

Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi. org/10.1136/bmjopen-2018-026138).

Received 31 August 2018 Revised 9 April 2019 Accepted 3 June 2019



@ Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by

For numbered affiliations see end of article.

#### **Correspondence to**

Dr Kavitha Subramaniam; kavitha.subramaniam@act. gov.au

#### **ABSTRACT**

Objective Post-colonoscopy colorectal cancers (PCCRCs) are recognised as a critical quality indicator. Benchmarking of PCCRC rate has been hampered by the strong influence of different definitions and methodologies. We adopted a rigorous methodology with high-detail individual data to determine PCCRC rates in a prospective cohort representing a single jurisdiction.

Setting We performed a cohort study of individuals who underwent colonoscopy between 2001 and 2008 at a single centre serving Australian Capital Territory (ACT) and enclaving New South Wales (NSW) region. These individuals were linked to subsequent colorectal cancer (CRC) diagnosis, within 5 years of a negative colonoscopy, through regional cancer registries and hospital records using probabilistic and deterministic record linkage. All cases were verified by pathology review. Predictors of PCCRCs were extracted.

Participants 7818 individuals had a colonoscopy in the cohort. Linkage to cancer registries detected 384 and 98 CRCs for notification dates of 2001-2013 (ACT) and 2001-2010 (NSW). A further 55 CRCs were identified from a search of electronic medical records using International Classification of Diseases-10 diagnosis codes. After verification and exclusions, 385/537 CRCs (58% male) were included.

Primary outcome measure PCCRC rates.

Results There were 15 PCCRCs in our cohort. The PCCRC incidence rate was 0.384/1000 person-years and the 5-year PCCRC risk was estimated as 0.192% (95% Cl 0.095 to 0.289). The index colonoscopy prior to PCCRC was more likely to show diverticulosis (p=0.017 for association, OR 3.56, p=0.014) and have poor bowel preparation (p=0.017 for association, OR 4.19, p=0.009). Conclusion In this population-based cohort study, the PCCRC incidence rate was 0.384/1000 person-years and the 5-year PCCRC risk was 0.192%. These data show the 'real world' accuracy of colonoscopy for CRC exclusion.

### INTRODUCTION

Colonoscopy is a recommended, preferred<sup>2</sup> and cost-effective screening modality for colorectal cancer (CRC) screening, with large observational demonstrating its effectiveness in reducing CRC incidence and mortality.3-5

### Strengths and limitations of this study

- ► The post-colonoscopy colorectal cancer (PCCRC) rate is a quality standard for colonoscopy, but may be markedly affected by the detail available in large datasets and the methodology used to link them.
- We used a prospective colonoscopy cohort linked to state cancer registries and the electronic medical record, a combination of probabilistic and deterministic linkage, and individual case review to derive the incidence of PCCRC.
- For benchmarking of PCCRC incidence, the cohort having index colonoscopy should be followed prospectively to ensure that the same population is assessed.
- Our approach was affected by a lag to registration in each cancer registry, limiting the cohort size with a minimum of 5 years' follow-up.

As with other forms of cancer screening, interval cancers, defined as CRC diagnosed after a screening or surveillance examination in which no cancer is detected, and before the date of the next recommended examination,<sup>6</sup> can occur. The Asia Pacific CRC Working Group have published recommendations on minimising interval cancers, recognising that interval cancers are a reflection of quality in colonoscopy practice. More recently, these events have been termed post-colonoscopy colorectal cancers (PCCRC), with the term interval cancer applied only to CRC arising within a defined period after a negative examination performed primarily for screening<sup>8</sup> and to differentiate those screened by another means. Expert opinion is that PCCRCs are potentially the most important markers of colonoscopy quality.

CRC is the second most common cancer in Australia. 10 As a result, even a low-frequency event such as PCCRC may yield a substantial number of cases. Internationally, the reported proportion of new CRC cases that are PCCRC varies widely, with rates up to 9%,<sup>11</sup> possibly because of differing methodology and quality. Methodological issues that impact most severely on rates of PCCRC have been addressed.<sup>6 12</sup>

One theoretically accurate way to determine PCCRC incidence is to measure its occurrence prospectively in a cohort of patients with colonoscopies negative for cancer. As detection of PCCRCs requires case matching between at least two datasets (such as Procedure or Claims Databases with a Cancer Registry), the relative completeness of data and quality of data linkage must also determine the apparent PCCRC incidence. The aim of this study was to accurately determine the PCCRC incidence in our geographically small but populous Australian jurisdiction, by exploiting the availability of three separate database sources—a Prospective Colonoscopy Database, a Hospital Records System and a Regional Cancer Registry. We sought to maximise one-to-one linkage and enable benchmarking in our jurisdiction and other regional jurisdictions.

# METHODS Study design

We performed a cohort study of individuals who underwent outpatient colonoscopy at the Canberra Hospital, Australia Capital Territory (ACT), Australia from 1 January 2001 to 31 December 2008, who were residents of the ACT at the time. We then identified individuals from this cohort subsequently diagnosed with CRC anywhere in ACT or New South Wales (NSW).

The ACT is a self-governing jurisdiction, with a population 390 800 and land area 2358 km², in the south-east of Australia. It is unique geographically in being enclaved by the larger and more populous NSW, with a population 7618 200 and 800 641 km². In addition, its major hospital, the Canberra Hospital, is also the major public hospital for the surrounding region, supporting a population of 540 000. While migration to other states occurs, the majority of residents leaving the ACT take up residence in NSW.<sup>13</sup>

The Canberra Hospital performs the majority of publicly funded colonoscopies performed in the ACT. All colonoscopies performed in the ACT are performed in accredited facilities. Approximately 30% of colonoscopies performed at the Canberra Hospital are completed by trainees under supervision by accredited endoscopists.

CRC data were available from the ACT Cancer Registry (2403 CRC records for the study period including minimum 5 years' follow-up, see table 1) and the NSW Cancer Registry (46200 CRC records for the period, table 1). The principal outcome measure of this study was PCCRCs, defined as CRC diagnosed anywhere in ACT or NSW within 5 years following a complete colonoscopy at Canberra Hospital. All colonoscopies performed at the Canberra Hospital between January 2001 and December 2008 that met inclusion criteria were used to report the incidence rate for PCCRC. In order to determine the incidence rate, we calculated the total person-years

**Table 1** Baseline characteristics of the cohort between 2001 and 2008 (9383 procedures in 7818 individuals)

Case characteristics	n	%
Sex		
Male	3610	46.2
Female	4208	53.8
Age (years)		
<40	1474	18.9
40–49	1585	20.3
50–59	1855	23.7
60–69	1542	19.7
>70	1360	17.4
Year of first colonoscopy	1	
2001	543	6.9
2002	1026	13.1
2003	1042	13.3
2004	929	11.9
2005	1061	13.6
2006	1048	13.4
2007	1047	13.4
2008	1122	14.3
Findings		
Normal	3702	39.5
Diverticular disease	1855	19.8
Haemorrhoids	1659	17.7
IBD/colitis	693	7.4
Colonic polyps	3113	33.2
First endoscopist		
Non-trainee	6165	65.7
Trainee	3209	34.2
Unknown	9	0.1

IBD, Inflammatory bowel disease.

of observation for each patient based on the interval between the date of the colonoscopy and the earliest of the following: the diagnosis date of CRC, date of death or the study census date (census date 10 February 2016). The incidence of PCCRC was computed as the observed number of PCCRC divided by the total person-years of observation.

# Patient and public involvement

We included all patients at our centre who had a complete colonoscopy during the recruitment period. The PCCRC incidence rate in a cohort undergoing colonoscopy is a fact that can be shared in the informed consent process with patients undergoing diagnostic colonoscopy.

### **Data sources**

All colonoscopies were reported through proprietary reporting software into a relational database. Patient identifiers for all procedures were populated from The Canberra Hospital patient administration system (CareSys, Advanced Health and Care, Ashford, UK, until September 2006; thereafter iPM, Healthcare Group, CSC, Banbury, UK).

The ACT Cancer Registry records all cases of cancer diagnosed in ACT residents since 1994 except basal cell carcinomas and squamous cell carcinomas. Pathology laboratories, hospitals and nursing homes in the ACT are required by law to notify the Registry of Cancer in people or human tissues. The NSW Central Cancer Registry maintains records of all cases of cancer diagnosed in NSW residents. The Centre for Health Record Linkage (CHeReL)<sup>14</sup> is a collaboration of partner institutions in ACT and NSW creating a record linkage infrastructure for the health sectors covered by the collaboration and links over 69 million records representing more than 9 million individuals. Record identifiers are included in the Master Linkage Key,<sup>15</sup> which has a false positive rate of 0.3%.<sup>14</sup>

To identify CRC cases missed by linkage between the colonoscopy database and the Cancer Registry, we also extracted CRC cases (adenocarcinomas only) for the same period from the Canberra Hospital electronic medical record by search of diagnostic codes C18.0, C18.1, C18.2, C18.3, C18.4, C18.5, C18.6, C18.7, C18.8, C18.9, C19, C20 from the International Classification of Diseases (ICD-10) code set. These cases were matched using the hospital patient identification number (PID) as the primary linkage and verified by first name, last name and date of birth. All matches were manually reviewed to ensure that the PID had not been previously assigned.

#### **Cohort at risk**

The cohort at risk included all patients who had a complete colonoscopy during the recruitment period between January 2001 and December 2008. The colonoscopy performed before the diagnosis of PCCRC was termed the index colonoscopy. A colonoscopy was considered complete if the proceduralist reported caecal landmarks or intubated the terminal ileum of an intact colon. Cases were excluded after manual record review if the colonoscopy was incomplete and the reason for incompletion was recorded. We included patients with elevated risk of CRC, such as Lynch syndrome, familial adenomatous polyposis and inflammatory bowel disease, in the cohort.

### **Case linkage**

The ACT and NSW Central Cancer Registry custodians provided record identification numbers for CRC cases for the recruitment period 1 January 2001 to 31 December 2008. ACT and NSW Cancer Registry records matching the CRC subset were then extracted from the Master Linkage Key for the periods of interest. This extract was linked to the colonoscopy records using probabilistic record linkage methods and *ChoiceMaker* software. <sup>16</sup>

#### **Data collection**

Data were extracted from colonoscopy reports, the electronic medical record and pathology reports. Case characteristics examined were age, gender and indication for colonoscopy. Colonoscopy characteristics examined included date of colonoscopy, the identity of first and (where applicable) second proceduralists, the sedation drug and dose, quality of bowel preparation, caecal and terminal ileal intubation rate (completion rate), the presence of postsurgical anatomy and all diagnoses made. The CRC site was classified as the proximal colon where this was the caecum to the transverse colon. The database was checked for internal validity by individual case review of all cases using the source medical record. Histopathological classification and staging were in accordance with the Cancer Staging Manual of the American Joint Committee on Cancer (AJCC). 17

# Determination of mismatch repair defective CRC including Lynch syndrome

Mismatch repair (MMR) deficiency was determined by immunohistochemistry for MLH1, MSH2, MSH6 and PMS2 proteins in the relevant formalin-fixed paraffin-embedded (FFPE) sections, using commercial antibodies (Ventana Medical Systems, Tucson, Arizona, USA). MMR expression status was available for all CRC diagnosed from 1 January 2004 in colon resection specimens but not where biopsy material only was available. BRAF V600E mutation status was determined retrospectively on MLH1-deficient cases from FFPE-extracted tumour DNA by single nucleotide primer extension assay. <sup>18</sup>

#### Statistical methods

Descriptive statistics were used to summarise the data. Fisher exact tests were carried out to explore the differences between groups. ORs for occurrence of PCCRC and non-PCCRC were calculated where indicated.

Kaplan-Meier estimates were used to estimate the 5-year PCCRC risk in this cohort. The time to PCCRC for patients diagnosed with PCCRC was calculated as the time from their last colonoscopy to the time of diagnosis. The time to last follow-up for all other patients was calculated as the time from their last colonoscopy to the census date 10 February 2016, when the linkage with the Cancer Registries was completed.

# **RESULTS**

### **Cohort characteristics**

There were 10 640 colonoscopy procedures carried out at the Canberra Hospital between January 2001 and December 2008. Of these, 3209 were performed by a trainee (first proceduralist) under supervision of an experienced second proceduralist. Sedation for the majority of procedures was administered by a nurse sedationist under direction of the endoscopist with a combination of fentanyl, midazolam and propofol (endoscopist-directed nurse administered propofol sedation). Specialist

anaesthetists provided sedation for patients classed as American Society of Anesthesiologists (ASA) Physical Status III.

Detailed case review of the entire patient group submitted for linkage was performed. A total of 465 procedures in 465 cases with a diagnosis of CRC at the initial colonoscopy were excluded from the cohort for the purposes of survival analysis. After line-by-line analysis of the database for internal agreement, agreement against the source medical record, and inclusion and exclusion criteria, another 792 procedures were excluded from further analysis. These included 166 incomplete procedures due to instrument looping, patient discomfort or bowel redundancy. A total of 101 procedures were excluded due to findings of a shortened postsurgical colon (ileostomy, rectal stump, neoterminal ileum or subtotal colectomy, but not anterior resection). A further 134 procedures were excluded due to incomplete colonoscopy arising from the pathology encountered, including obstructing cancer or other colonic stricture. In 133 cases, an elective short colonoscopy was performed and these cases were also excluded. Another 210 procedures were excluded due to bowel preparation classed by the reporting proceduralist as 'poor', 'unsatisfactory' or 'inadequate' that precluded caecal intubation. Cases were not excluded for poor bowel preparation if caecal intubation was achieved. A further 36 procedures were excluded where difficulties with patient sedation caused incompletion.

Characteristics of the overall cohort at risk (7818 individuals; 9383 procedures) are summarised in table 1. The average age at the time of the index colonoscopy was 53.6 years (SD 15.7) and 53.3% of the cohort was female. Follow-up ranged from 5 to 13 years (median follow-up, 7.9 years). At colonoscopy, 39.5% (n=3702) were reported as normal, while 33.2% (n=3113) involved polypectomy. Other findings included diverticulosis in 19.8%, haemorrhoids in 17.7% and inflammatory bowel disease in 7.4%.

# Results of data linkage

The datasets for linkage are shown in table 2. We submitted a total of 10 854 individuals who had one or more colonoscopy at our centre during the period 1 January 2001 to 31 December 2010 for linkage. This group contained individuals who were later excluded from the final cohort after detailed case review. Total CRC notifications to the ACT Cancer Registry and NSW Central Cancer Registry were 2403 records (notification date 1 January 2001 to 31 December 2013) and 46 200 records (notification date 1 January 2001 to 31 December 2010), respectively (table 2). Linkage from submitted records to these ACT and NSW CRC notifications detected 384 and 98 CRCs for notification dates of 2001–2013 (ACT) and 2001–2010 (NSW; online supplementary table 1).

Next, cases retrieved from a search of The Canberra Hospital electronic medical records using the relevant ICD-10 diagnosis codes were matched to the colonoscopy database according to a complete match of PID,

Table 2 Datasets used in this study for data linkage						
Data source	Description	Number				
Canberra Hospital Gastroenterology and Hepatology Unit colonoscopy cohort	Procedure dates: 1 Jan 2001 to 31 Dec 2013 Subset: Procedure dates 1 Jan 2001 to 31 Dec 2010	19803 records 13746 records				
	ACT Cancer notifications: Notification date: 1 Jan 2001 to 31 Dec 2013	19931 records				
CCR)	Subset: Colorectal cancer notifications: Notification date: 1 Jan 2001 to 31 Dec 2013	2403 records				
NSW Cancer Registry colorectal cancer subset	NSW cancer notifications: Notification date: 1 Jan 2001 to 31 Dec 2010	383213 records				
	Subset: Colorectal cancer notifications: Notification date: 1 Jan 2001 to 31 Dec 2010	46200 records				

ACT, Australian Capital Territory; NSW, New South Wales.

first name, last name and date of birth. A further 55 CRCs were linked that had not been returned from linkage to the ACT or NSW Cancer Registries.

From the total of 537 CRCs, 384 returned from the ACT Cancer Registry, 98 from the NSW Cancer Registry and 55 from the Canberra Hospital electronic medical record, 385 were included in this study after excluding cases that did not have a colonoscopy during the recruitment period 2001–2008, those with a diagnosis of malignant polyp, carcinoid tumour, metastatic malignancy of unknown primary, CRC diagnosed outside NSW or ACT, and cases that could not be verified by pathological or clinical records. A summary of the results of linkage in the study is presented in figure 1.

# Reasons for linkage failure

There were 55 cases not returned from linkage to the cancer registries, which were subsequently returned from the Canberra Hospital electronic medical record. Of these 55 cases, 21 cases were included in this study after case review. The principal reasons for linkage failure were migration from the ACT to states other than NSW (n=15), missing identifiers (n=5) and one-to-many links (n=1).

### **PCCRC** incidence

Of the 385 CRC cases returned from database linkage, 15 cases had a colonoscopy at our centre during the recruitment period and were diagnosed 0–60 months after the index colonoscopy, and these cases were defined as PCCRC. There were an additional 19 CRC that met some but not all criteria for PCCRC. In one case, case review revealed that the index colonoscopy failed to reach the

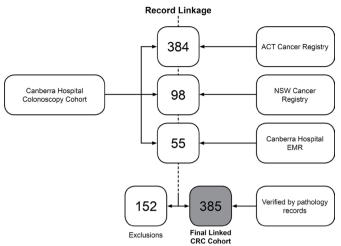


Figure 1 Results of data linkage. Primary linkage between the colonoscopy cohort and the ACT and NSW Cancer Registries by the Centre for Health Record Linkage (CHeReL) yielded 384 and 98 CRCs for notification dates of 2001–2013 (ACT) and 2001–2010 (NSW). An independent match of the colonoscopy database and The Canberra Hospital electronic medical record (EMR) matched 55 CRC to that had not been returned from linkage to the Cancer Registries. From this total of 537 CRCs, 152 were excluded where first colonoscopy fell outside the recruitment period, the CRC was diagnosed outside NSW/ACT or diagnosis of malignant polyp or carcinoid tumour or metastatic malignancy of unknown primary and cases that could not be verified. ACT, Australian Capital Territory; CRC, colorectal cancer; NSW, New South Wales.

caecum. In one case, the index colonoscopy occurred prior to the recruitment period. In seven cases, the diagnosis of CRC was made more than 5 years but less than 10 years after the previous index colonoscopy. <sup>19</sup> There were 10 cases with multiple exclusion criteria.

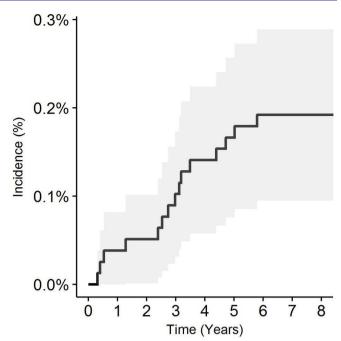
The total exposure period of  $81\,879$  person-years in our cohort resulted in a CRC incidence rate of 4.70/1000 person-years. For PCCRC, the first 5 years following a colonoscopy contributed to total exposure and the PCCRC incidence was 0.384/1000 person-years.

Kaplan-Meier estimates were used to estimate the 5-year PCCRC risk in this cohort. The 5-year PCCRC risk was estimated as 0.192% (95% CI 0.095 to 0.289) (figure 2).

#### Factors associated with increased risk of PCCRC

Univariate analysis was performed to identify characteristics associated with PCCRC cases (table 3). The presence of diverticulosis at index colonoscopy (p=0.017, Fisher's exact test) was associated with PCCRC (table 3). The OR for the presence of diverticulosis and subsequent diagnosis of PCCRC was 3.56 (95% CI 1.29–9.83, p=0.014). The number needed to treat (NNT) for the presence of diverticulosis and subsequent PCCRC was 333.

The presence of poor bowel preparation was associated with PCCRC (p=0.017, Fisher's exact test). OR for the presence of poor bowel preparation and subsequent diagnosis of PCCRC was 4.19 (95% CI 1.43 to 12.3, p=0.009).



**Figure 2** Kaplan-Meier incidence plot (solid line) and 25% Cls (shaded area) of post-colonoscopy colorectal cancer arising in the cohort. The solid line represents patients having colonoscopy at the Canberra Hospital during 2001–2008. Each increment in the Kaplan-Meier plot is an occurrence of post-colonoscopy colorectal cancer.

The NNT for the presence of poor bowel preparation and subsequent PCCRC was 250.

Trainee endoscopists were listed as the first endoscopists of index colonoscopies associated with 53.3% (8/15) PCCRC compared with 46.7% of the rest of the cohort, where the first listed endoscopists was not a trainee (p=0.170, Fisher's exact test).

# PCCRC characteristics compared with incident colorectal cancers

There were a total of 385 CRC returned from linkage and comprised 258 incident CRC diagnosed at the index colonoscopy. A further 93 CRC cases underwent index

**Table 3** Association between index colonoscopy characteristics and occurrence of PCCRC

Characteristics	OR*	95% CI	P value	Number needed to treat
Diverticulosis at colonoscopy and PCCRC	3.56†	1.29 to 9.83	0.0143	333
'Poor' bowel preparation at colonoscopy and PCCRC	4.19*	1.43 to 12.3	0.0091	250

<sup>\*</sup>Compared with group with 'good' bowel preparation.

<sup>†</sup>Compared with group with no finding of diverticulosis.

Bold values highlight significance.

PCCRC, post-colonoscopy colorectal cancer.

**Table 4** Characteristics of PCCRCs compared with incident CRC (n=273)

Characteristic	PCCRC cases, N (%)	Incident CRC, n (%)	P value			
Gender						
Male	5 (33.3)	153 (59.3)				
Female	10 (66.7)	105 (40.7)	0.048			
Age (years)						
≤60	5 (33.3)	74 (28.7)				
>60	10 (66.7)	184 (71.3)	0.699			
Unknown		5 (1.30)				
Indication						
Diagnostic	7 (46.7)	215 (83.3)				
Surveillance	6 (40)	24 (9.3)	0.0005			
Screening	2 (13.3)	16 (6.2)				
Unknown	0 (0)	3 (0.12)				
Diverticulosis						
Yes	6 (40)	43 (16.7)				
No	9 (60)	215 (83.3)	0.022			
First endoscopist	t					
Non-trainee	7 (46.7)	176 (68.2)				
Trainee	8 (53.3)	82 (31.8)	0.084			
Bowel preparation						
Good	5 (33.3)	71 (27.5)				
Satisfactory	5 (33.3)	123 (47.7)				
Poor	5 (33.3)	62 (24)	0.526			
Unknown	0 (0)	2 (0.78)				

CRC, colorectal cancer; PCCRC, post-colonoscopy colorectal cancer.

Bold values highlight significance.

colonoscopy after the diagnosis of CRC. There were 34 provisional PCCRC diagnosed after the index colonoscopy and before the study census date. Of these 34 provisional PCCRC, 7 cases were diagnosed 5 years after the colonoscopy and 12 cases did not meet inclusion criteria. A total of 15 cases meeting our criteria for PCCRC were further analysed.

In the overall cohort, 66.7% of cases with PCCRCs were females compared with 40.7% of incident CRCs diagnosed (p=0.048, table 4). The mean age of PCCRC cases was 65 years and that of incident CRCs was 68 years. Clinicopathological characteristics of all CRC are summarised in online supplementary table 2. Among the 370 CRC that were not PCCRC, there was one case with mutation-proven Lynch syndrome and one case with mutation-proven familial adenomatous polyposis. There were two cases of Lynch syndrome in the 15 PCCRC cases; however, these were ascertained after the diagnosis of CRC.

CRC were located in the proximal colon in  $8\ (53.3\%)$  PCCRC compared with  $80\ (31\%)$  of incident CRC

(p=0.48). There was no association of PCCRC with histological subtype (p=0.17) or with CRC stage (p=0.57; online supplementary table 2). Of 15 PCCRC, 7 were detected in individuals who had prior polypectomy; in 2 of these cases, a polyp was removed in the same region of the colon.

One hundred fifty-one CRC cases overall had available MMR expression status during the period (online supplementary table 3). There were no statistically significant differences between PCCRC and incident CRC with respect to abnormal MMR status (30% vs 20.6%; p=0.48). BRAF V600E mutation status was assessed in both cases of MLH1-negative PCCRC and 22/29 of other CRCs with absent MLH1 expression. The BRAF V600E mutation was detected in 50% (1/2) of the MLH1-negative PCCRCs and in 82% (18/22) of the MLH1-negative CRCs that were not PCCRCs.

#### DISCUSSION

Since a comprehensive study reported in 2007 the 'miss rates' for CRC in subjects having colonoscopy,<sup>20</sup> it has been accepted that PCCRC comprise a measurable and significant proportion of CRC cases. A recent consensus statement holds that PCCRC is the preferred term for cancers appearing after a colonoscopy in which no cancer is diagnosed.<sup>19</sup> As a quality assurance term, PCCRC refers to CRC detected after any such complete prior colonoscopy. The term 'interval cancer' was recommended for CRC (or other cancer) arising before the next recommended screening or surveillance procedure.

In our study, we followed a cohort of 7818 individuals who underwent 10640 colonoscopy procedures at our centre over the period 1 January 2001 to 31 December 2008, for a minimum of 5 years following the index colonoscopy. We then performed linkage for the detection of subsequent CRC. In this prospective, population-based cohort study, 15 individuals developed PCCRC, resulting in an incidence rate of 0.384/1000 person-years. The 5-year probability of PCCRC was 0.192%.

A significant factor that arose in the study was the number of index colonoscopies that were excluded because of incompletion, previous surgery, pathology or technical factors. The effect of exclusions was to reduce the overall cohort size and thereby increase the apparent incidence rate. Making these exclusions required line-by-line verification of the endoscopy database and access to source documents. This process is unavailable to studies that use larger volume retrospective data sources.

We also audited case linkage by a search of the electronic medical record for unlinked cases of CRC. We found 21 additional cases of CRC (5% of the total), which had escaped the linkage process, predominantly because the individual had migrated from the jurisdiction during the follow-up period. These cases did however present with CRC to our centre within the jurisdiction. We suspect these 21 cases to be a minority of the occasions of linkage failure in this study. While our study benefited from the

Author, Year, (reference)	Period & Country	Design	Definition of interval CRC	Data Sources	Primary Data set	Methods of Record linkage	Interval CRC rate
Baxter 2011	2000 -2005 Ontario, Canada	Retrospective cohort, administrative database record linkage	CRC diagnosis 6-36 months after colonoscopy	Ontario Health Insurance Plan (OHIP) Database (physician claims database to identify colonoscopy) Canadian Institute for Health Information (ICD-9 diagnosis codes) Ontario Cancer Registry	Ontario Cancer Registry	OHIP number is used to match Ontario Cancer Registry to other records.	9%
Brenner 2012	2003-2007, Rhine-Neckar region, Germany	Prospective, case- control, 22 hospitals participated	CRC diagnosis 1-36 months after negative colonoscopy	Screen-detected CRC database     Incident cancer database: mostly surgical patients     Group 1 and 2 interviewed in hospital. For any history and date of previous colonoscopy, some validation done with physician records	Screen-detected and incident CRC database	N/A	1.8%
Brenner 2012	2003-2007, Rhine-Neckar region, Germany	Prospective, case- control, 22 hospitals participated	CRC diagnosis 1-36 months after colonoscopy and polypectomy	Screen-detected CRC database     Incident cancer database: mostly surgical patients     Group 1 and 2 interviewed in hospital. For any history and date of previous colonoscopy, some validation done with physician records	Screen-detected and incident CRC database	N/A	2.3%
Bressler 2007	1997-2002 Ontario, Canada	Retrospective cohort, administrative database record linkage	CRC diagnosis 6-36 months after colonoscopy	Ontario Health Insurance Plan Claims Database Canadian Institute for Health Information Ontario Cancer Registry	Ontario Cancer Registry	Health Care Number from Ontario Cancer Registry used to match OHIP cases	3.4%
Cooper 2011	1994-2005, USA	Retrospective cohort, administrative database record linkage; Age limited to >69	CRC diagnosis 6-36 months after colonoscopy	<b>SEER/Medicare</b> database (contains separate data from files of providers, outpatients and inpatient facilities)	SEER	Shared identifier	7.2%
Erichsen 2013	2000-2009, Denmark	Retrospective cohort; population-based with record linkage	CRC diagnosis 12-60 months after colonoscopy	Danish National Registry of Patients (tracks hospitalisations and outpatient visits) to get colonoscopy exams Danish Cancer Registry	Danish Cancer Registry	Same Civil Registration Number (10 digit unique ID number) in both data sets	2.8%
Farrar 2006	1991-2004, Minneapolis, Minnesota, USA	Retrospective cohort; population-based; Veterans only	CRC diagnosis 60 months after complete colonoscopy	Minneapolis Veterans Affairs Medical Centre Cancer Registry Minneapolis Veterans Affairs Medical Centre Medical Records Department	Minneapolis Veterans Affairs Medical Centre Cancer Registry	Manual lookup of medical records	5.4%
Ferrandez 2010	2000-2005, Hospital Clinic Lozano Blesa Zaragoza, Spain	Retrospective cohort	CRC diagnosis 36 months after colonoscopy	Colonoscopy reports at a single centre	Colonoscopy reports (Electronically generated reports not used; open boxes generate nonstandardized reports – some data lacking)	N/A	7% (27/386)
Horiuchi 2011	2005-2009, Showa Inan General Hospital, Japan	Retrospective cohort	CRC diagnosis 60 months after negative colonoscopy	Endoscopy database at a single centre (2500 colonoscopies/year performed)	Endoscopy database	N/A	4.3% (9/211)
Huang 2012	May 1976 - June 2008, Nanfang Hospital, Guangzhou and Affiliated Hospital of Inner Mongolia Medical College, Huhhot, China	Retrospective cohort	CRC diagnosis 60 months after complete colonoscopy and polypectomy	Patient record system and medical records at two hospitals	patient record system and medical records	N/A	Among 1,794 patients,14 had interval CRC before scheduled colonoscopy
Kaminski 2010	Oct 2000- Dec 2004, Poland	Retrospective cohort; population-based; 40-66 years old	CRC diagnosis before scheduled colonoscopy	National Colorectal Cancer Screening Program Database (40-66 yrs old average risk) Regional and National Cancer Registries	National Colorectal Cancer Screening Program Database	Used personal identifiers from CRC Screening Database to search Cancer Registries	42 interval CRC during 188,788 person-year
Le Clercq 2014	2001-2010; South Limburg, the Netherland	Retrospective cohort; population-based with record linkage; 3 hospitals covers for 650000 inhabitants	CRC diagnosis 12-60 months after colonoscopy	PALGA Database (pathology database); Medical record review at each of the 3 hospitals	PALGA Database verified by Cancer Registry	N/A	2.9%
Samadder 2014	1995-2009; Utah, United States	Retrospective cohort; population-based with record linkage	CRC diagnosis 6-60 months after colonoscopy	Univ of Utah Healthcare and Intermountain Healthcare EMR procedure codes Utah Cancer Registry	Univ of Utah Healthcare and Intermountain Healthcare EMR procedure codes	Probabilistic record linkage	3.4% (6-36m) 6% (6-60m)
Singh 2010	1992-2008; Manitoba, Canada;	Retrospective cohort; population-based with record linkage; 50-80 years old	CRC diagnosis 6-36 months after colonoscopy	Manitoba Health Databases: (1) MH Physicians Billing Claim Database (to identify colonoscopy; colonoscopy reports not visualised) (2) MH Population Registry (migration status of individuals) (3) MH Hospital Discharge Database (hosp discharges)	Manitoba Health Databases	Deterministic using Patient Health Identification Number to Manitoba Health Cancer Registry	7.9%

Figure 3 Interval cancer rates expressed as a proportion of all colorectal cancer (CRC) from published studies

Table 5 Incidence rates per 1000 patient-years of PCCRC from eight published articles during 2000–2017 and the present study

Study	Total number of subjects	Follow-up (years)	Cases of PCCRC	Incidence per 1000 person-years
Alberts et al <sup>33</sup>	1303	35 months (median)	9	2.4
Citarda et al <sup>34</sup>	1693	10.5 (mean)	6	0.4
Robertson et al <sup>35</sup>	2915	3.7 (mean)	19	1.7
Pabby et al <sup>36</sup>	2079	2.79 (mean)	13	2.2
Lieberman et al <sup>37</sup>	1171	5.5 (mean)		1.7
Leung et al <sup>38</sup>	1297	10 (median)	9	1.2
Belderbos et al <sup>39</sup>	107744	5.1 (mean)	1031	1.88
Samadder et al <sup>40</sup>	131 349		171	0.55
Present study	7818	10.52 (mean)	15	0.38

PCCRC, post-colonoscopy colorectal cancer.

peculiarities of the ACT's geography, we have no information as to whether unlinked cases of CRC are more or less likely to be PCCRC.

A remarkable feature of interval CRC and PCCRC studies is the variation between PCCRC rates expressed as a proportion of all CRC, in different studies, from as low as  $0.8\%^{21}$  to as high as  $9\%^{11}$ . Among studies that used a single identifier to link two databases (figure 3), 11 20 22-32 the rate of PCCRC ranged from 2.8% to 9%.  $^{11\ 20\ 23\ 25\ 28}$ Two studies reporting from the same jurisdiction yielded different PCCRC rates of 3.4% and 9%. 11 20 It follows that these variations are due less to differences in CRC biology or technical excellence in colonoscopy than to methodological issues. In some retrospective studies, <sup>26</sup> <sup>27</sup> where the PCCRC rate appears to be lower (1.8%-2.3%), ascertainment of a prior colonoscopy was based on self-reporting by participants with recently diagnosed CRC. One study reporting a PCCRC rate of 2.9%<sup>22</sup> was performed by retrieving regional CRC records from a national pathology database and matching these to records from three regional hospitals to obtain a history of colonoscopy in the 5 years prior to diagnosis. The method of data linkage was not reported.

Alternatively, PCCRC have also been reported in cohort studies with the incidence rates expressed as a survival function; typically, 1.7–2.4 cancers per 1000 person-years of observation (table 5). Most have been interval CRC studies including only subjects with adenoma at index colonoscopy, which would be expected to yield a higher incidence rate of PCCRC than we found in our cohort.

Since we were limited to Cancer Registry notifications to the end of 2013 (for ACT cases), and the end of 2010 (for the smaller cohort of NSW cases), the follow-up period in the study population was 5–13 years, with a median follow-up period of 7.9 years. Cancer Registry data (online supplementary table 1) were available until 32 months prior to the linkage date (ACT Registry) and 68 months prior to the linkage date (NSW Registry). This

limited the size of the cohort that could be assessed for PCCRC with a follow-up of 5 years.

We report that linkage failure can be detected and should be addressed in PCCRC studies. We identified 21 cases of CRC that were not returned from data linkage, although none of these cases were PCCRC. Possible causes of linkage failure include missing or ambiguous identifiers, returning 'one-to-many' outputs. In order to identify linkage failure, we also matched CRC cases from our hospital medical record system.

We confirm the finding of others 20 23 25 28 that a characteristic strongly associated with PCCRC was a finding of diverticular disease at index colonoscopy. On the contrary, the OR for patients with diverticular disease and a subsequent diagnosis of PCCRC was 3.56. The NNT for the presence of diverticulosis and subsequent PCCRC was 333. In order to achieve a halving of the PCCRC ratio, for example, through 3 yearly surveillance colonoscopy, 666 patients would be required to undergo enhanced screening for each PCCRC avoided. Our study also found that poor bowel preparation increased the risk of PCCRC with an OR of 4.2.

We also confirmed the reported associations of PCCRC with female gender, <sup>11</sup> <sup>19</sup> <sup>22</sup> <sup>27</sup> but not with age, <sup>20</sup> <sup>23–25</sup> proximal tumour location <sup>20</sup> <sup>24–26</sup> <sup>28</sup> or family history of CRC. <sup>22</sup> <sup>24</sup> There was a non-significant association between index colonoscopies performed by a trainee and subsequent PCCRC. These negative findings may be due to type II error

The reported higher frequency of right-sided PCCRC may arise from these CRC being more biologically aggressive or alternatively from right-sided precursor lesions being more likely to be undetected or incompletely removed at prior examinations. The former explanation is supported by the observation of a higher frequency of abnormal MMR expression in PCCRC. However, recently Soong and colleagues performed analysis of cancer gene mutations and copy number variation by

targeted exon capture and deep sequencing in interval CRC detected at a single institution. When interval CRC were compared with incident CRC matched to age, gender and tumour location, there were no appreciable differences in any classifying molecular alteration in interval CRCs, pointing to a similar biological behaviour of interval and incident CRC. Soong *et al* concluded that interval PCCRC were therefore likely to result from missed or recurrent lesions. We did not find a significant difference in abnormal expression of MMR proteins between PCCRC and incident cancers. We found 2 of 15 PCCRC were detected in individuals who had prior polypectomy in the same region of the colon.

We note that our study design optimises detection of PCCRC arising in our cohort. We did not detect all PCCRC occurring during the period. Some CRC diagnosed during the period of this study may have been PCCRC if an index colonoscopy had been performed elsewhere within the previous 10 years. All these cases were excluded from our cohort but were used as a comparator group for analysis of factors associated with PCCRC. The degree to which this influences our comparisons was not ascertained.

Our institution performs the majority of publicly funded colonoscopies performed in the region. We have no data comparing our patient population with the overall regional population. The ACT population is comparable to the broader Australian population, but to a significant degree is more affluent and more likely to hold tertiary education qualifications, and is somewhat less linguistically and culturally diverse. With these caveats, we believe our results to broadly reflect the Australian population.

One factor distinguishing our facility from other local facilities was that trainees completed 34% of index colonoscopies in the cohort. There was a non-significant trend for these cases to be associated with PCCRC.

We conclude that in the Australian setting, in a cohort undergoing colonoscopy, using probabilistic record linkage and individually curated cases: the 5-year incidence rate for PCCRC was estimated as 0.4 per 1000 person-years. These data show the 'real world' accuracy of colonoscopy for CRC exclusion. These data should be shared with patients in the informed consent process before screening and diagnostic colonoscopy.

#### **Author affiliations**

<sup>1</sup>Gastroenterology and Hepatology Unit, Canberra Hospital, Canberra, Australian Capital Territory, Australia

<sup>2</sup>Australian National University Medical School, Canberra, Australian Capital Territory, Australia

<sup>3</sup>Cancer Research, Canberra Hospital, Canberra, Australian Capital Territory, Australia

<sup>4</sup>Statistical Consulting Unit, Australian National University, Canberra, Australia
 <sup>5</sup>Department of Anatomical Pathology, ACT Pathology, Canberra Hospital, Canberra, Australian Capital Territory, Australia

**Acknowledgements** The authors wish to acknowledge the contributions of the Centre for Health Record Linkage.

Contributors KS, DT and PWA performed study concept and design, analysis and interpretation of the data, drafting of the manuscript and critical revision of the manuscript for important intellectual content. TN performed study concept and design and critical revision of the manuscript for important intellectual content. MF performed analysis and interpretation of the data and critical revision of the manuscript for important intellectual content.

**Funding** The project was funded entirely by the Canberra Hospital Private Practice Trust Fund.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Ethical approval was obtained from the ACT Health Human Research Ethics Committee and the NSW Population and Health Services Research Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

**Data sharing statement** Additional data is available by emailing the corresponding author.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

# **REFERENCES**

- Rex DK, Johnson DA, Anderson JC, et al. American College of Gastroenterology guidelines for colorectal cancer screening 2009 [corrected]. Am J Gastroenterol 2009;104:739–50.
- Marbet UA, Bauerfeind P, Brunner J, et al. Colonoscopy is the preferred colorectal cancer screening method in a population-based program. Endoscopy 2008;40:650–5.
- Winawer SJ, Zauber AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med 1993;329:1977–81.
- Singh H, Turner D, Xue L, et al. Risk of developing colorectal cancer following a negative colonoscopy examination: evidence for a 10year interval between colonoscopies. JAMA 2006;295:2366–73.
- Baxter NN, Warren JL, Barrett MJ, et al. Association between colonoscopy and colorectal cancer mortality in a US cohort according to site of cancer and colonoscopist specialty. J Clin Oncol 2012;30:2664–9.
- Sanduleanu S, le Clercq CM, Dekker E, et al. Expert Working Group on 'Right-sided lesions and interval cancers', Colorectal Cancer Screening Committee, World Endoscopy Organization. Definition and taxonomy of interval colorectal cancers: a proposal for standardising nomenclature. Gut 2015;64:1257–67.
- Sung JJ, Ng SC, Chan FK, et al. An updated Asia Pacific Consensus Recommendations on colorectal cancer screening. Gut 2015;64:121–32.
- Shaukat A, Rector TS, Church TR, et al. Longer withdrawal time is associated with a reduced incidence of interval cancer after screening colonoscopy. Gastroenterology 2015;149:952–7.
- Rees CJ, Bevan R, Zimmermann-Fraedrich K, et al. Expert opinions and scientific evidence for colonoscopy key performance indicators. Gut 2016:65:2045–60.
- Australian Institute of Health and Welfare. Cancer incidence projections: Australia, 2011 to 2020. Cancer Series no. 66. Cat. No. CAN 62. Canberra: AIHW, 2012.
- Baxter NN, Sutradhar R, Forbes SS, et al. Analysis of administrative data finds endoscopist quality measures associated with postcolonoscopy colorectal cancer. Gastroenterology 2011;140:65–72.
- Morris EJ, Rutter MD, Finan PJ, et al. Post-colonoscopy colorectal cancer (PCCRC) rates vary considerably depending on the method used to calculate them: a retrospective observational populationbased study of PCCRC in the English National Health Service. Gut 2015;64:1248–56.
- Australian Bureau of Statistics. Labour force, Australia, 'Interstate Migration by States and Territories of Arrival and Departure by Sex ', time series spreadsheet. 2016 http://stat.abs.gov.au//Index.aspx? QueryId=1289# (viewed 23 Jul 2016).
- Centre for Health Record Linkage. 2006 http://www.www.cherel.org. au (accessed 03 Mar 2014).

- Roos LL, Wajda A. Record linkage strategies. Part I: Estimating information and evaluating approaches. *Methods Inf Med* 1991;30:117–23.
- Borthwick A, Buechi M, Goldberg A. Key Concepts in the ChoiceMaker 2 Record Matching System. Proceedings of the KDD-2003 Workshop on Data Cleaning, Record Linkage, and Object Consolidation. Washington, DC, 2003.
- AJCC. Colon and rectum. In: Edge SB, Byrd DR, Compton CC, eds. AJCC Cancer Staging Manual. 7th edn. New York, NY: Springer, 2010:143–64.
- Caruso M, Fung KY, Moore J, et al. Claudin-1 expression is elevated in colorectal cancer precursor lesions harboring the BRAF V600E Mutation. Transl Oncol 2014;7:456–63.
- Rutter MD, Beintaris I, Valori R, et al. World endoscopy organization consensus statements on post-colonoscopy and post-imaging colorectal cancer. Gastroenterology 2018;155:909–25.
- Bressler B, Paszat LF, Chen Z, et al. Rates of new or missed colorectal cancers after colonoscopy and their risk factors: a population-based analysis. Gastroenterology 2007;132:96–102.
- Huang Y, Gong W, Su B, et al. Risk and cause of interval colorectal cancer after colonoscopic polypectomy. *Digestion* 2012;86:148–54.
- le Clercq CM, Bouwens MW, Rondagh EJ, et al. Postcolonoscopy colorectal cancers are preventable: a population-based study. Gut 2014;63:957–63.
- Erichsen R, Baron JA, Stoffel EM, et al. Characteristics and survival of interval and sporadic colorectal cancer patients: a nationwide population-based cohort study. Am J Gastroenterol 2013;108:1332–40.
- Samadder NJ, Curtin K, Tuohy TM, et al. Characteristics of missed or interval colorectal cancer and patient survival: a population-based study. Gastroenterology 2014;146:950–60.
- Singh H, Nugent Z, Demers AA, et al. Rate and predictors of early/missed colorectal cancers after colonoscopy in Manitoba: a population-based study. Am J Gastroenterol 2010;105:2588–96.
- Brenner H, Chang-Claude J, Seiler CM, et al. Interval cancers after negative colonoscopy: population-based case-control study. Gut 2012;61:1576–82.
- Brenner H, Chang-Claude J, Rickert A, et al. Risk of colorectal cancer after detection and removal of adenomas at colonoscopy: population-based case-control study. J Clin Oncol 2012;30:2969–76.
- 28. Cooper GS, Xu F, Barnholtz Sloan JŚ, *et al.* Prevalence and predictors of interval colorectal cancers in medicare beneficiaries. *Cancer* 2012;118:3044–52.
- Farrar WD, Sawhney MS, Nelson DB, et al. Colorectal cancers found after a complete colonoscopy. Clin Gastroenterol Hepatol 2006;4:1259–64.

- Ferrández A, Navarro M, Díez M, et al. Risk factors for advanced lesions undetected at prior colonoscopy: not always poor preparation. *Endoscopy* 2010;42:1071–6.
- Horiuchi A, Nakayama Y, Kajiyama M, et al. Invasive colorectal cancer within 5 years of negative colonoscopy in a Japanese population. Colorectal Dis 2012;14:1090–4.
- Kaminski MF, Regula J, Kraszewska E, et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med 2010;362:1795–803.
- Alberts DS, Martínez ME, Roe DJ, et al. Lack of effect of a highfiber cereal supplement on the recurrence of colorectal adenomas. Phoenix Colon Cancer Prevention Physicians' Network. N Engl J Med 2000;342:1156–62.
- 34. Citarda F, Tomaselli G, Capocaccia R, et al. Efficacy in standard clinical practice of colonoscopic polypectomy in reducing colorectal cancer incidence. *Gut* 2001;48:812–5.
- Robertson DJ, Greenberg ER, Beach M, et al. Colorectal cancer in patients under close colonoscopic surveillance. Gastroenterology 2005;129:34–41.
- Pabby A, Schoen RE, Weissfeld JL, et al. Analysis of colorectal cancer occurrence during surveillance colonoscopy in the dietary Polyp Prevention Trial. Gastrointest Endosc 2005;61:385–91.
- Lieberman DA, Weiss DG, Harford WV, et al. Five-year colon surveillance after screening colonoscopy. Gastroenterology 2007;133:1077–85.
- Leung K, Pinsky P, Laiyemo AO, et al. Ongoing colorectal cancer risk despite surveillance colonoscopy: the Polyp Prevention Trial Continued Follow-up Study. Gastrointest Endosc 2010;71:111–7.
- Belderbos TD, Pullens HJ, Leenders M, et al. Risk of postcolonoscopy colorectal cancer due to incomplete adenoma resection: A nationwide, population-based cohort study. *United European Gastroenterol J* 2017;5:440–7.
- Samadder NJ, Pappas L, Boucherr KM, et al. Long-term colorectal cancer incidence after negative colonoscopy in the state of Utah: The Effect of Family History. Am J Gastroenterol 2017;112:1439–47.
- Singh S, Singh PP, Murad MH, et al. Prevalence, risk factors, and outcomes of interval colorectal cancers: a systematic review and meta-analysis. Am J Gastroenterol 2014;109:1375–89.
- Soong TR, Nayor J, Stachler MD, et al. Clinicopathologic and genetic characteristics of interval colorectal carcinomas favor origin from missed or incompletely excised precursors. Mod Pathol 2019;32:666–74.
- Australian Bureau of Statistics. QuickStats. 2018 https://quickstats.censusdata.abs.gov.au/census\_services/getproduct/census/2016/quickstat/CED801