







## CRCBaSe: a Swedish register-based resource for colorectal adenocarcinoma research

Caroline E. Weibull<sup>a</sup> , Sol Erika Boman<sup>a</sup>, Bengt Glimelius<sup>b</sup> , Ingvar Syk<sup>c</sup> , Peter Matthiessen<sup>d</sup> , Karin E. Smedby<sup>a</sup>, Caroline Nordenvall<sup>e,f,\*</sup>  and Anna Martling<sup>e,f,\*</sup> 

<sup>a</sup>Department of Medicine Solna, Clinical Epidemiology Division, Karolinska Institutet, Stockholm, Sweden; <sup>b</sup>Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; <sup>c</sup>Department of Surgery, Skåne University Hospital, Malmö, Sweden; <sup>d</sup>Department of Surgery, Faculty of Medicine and Health, Örebro University, Örebro, Sweden; <sup>e</sup>Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; <sup>f</sup>Department of Pelvic Cancer, GI Oncology and Colorectal Surgery Unit, Karolinska University Hospital, Stockholm, Sweden

### ABSTRACT

**Objectives:** To facilitate high-quality register-based research on colorectal cancer (CRC) in Sweden by constructing a database consisting of CRC patients, matched comparators, and relatives.

**Material and methods:** Patients with adenocarcinoma in the colon and/or rectum were identified in the Swedish Colorectal Cancer Register, a nationwide quality-of-care register. For each patient, six comparators from the general population were matched on birth year, sex, year of CRC diagnosis, and county. Comparators were free from CRC at the time of matching, but could later become cases. For both patients and comparators, first-degree relatives (parents, siblings, and children) were identified. Information from nationwide population-based registers was retrieved and linked to each individual in the database using the personal identification number unique to all Swedish residents.

**Results:** A total of 76,831 CRC patients diagnosed between 1995 and 2016 were identified (51% colon, 49% rectal; before 2007 only rectal cancer patients were included). Among all patients, 37% were stage I–II, 22% stage III, and 22% stage IV. The median follow-up time was 11.9 years (inter-quartile range, IQR: 8.6–15.3). Together with comparators and relatives, the database contains 2,413,139 individuals with information on demographics, dates and causes of death, in- and outpatient healthcare records, cancer diagnoses, prescribed and dispensed drugs, childbirths (among women), and social security information (such as sick leave and early retirement).

**Conclusion:** The Colorectal Cancer Database Sweden (CRCBaSe) is a large and unique register-based data research platform, which opens up for clinically important, large epidemiological studies with innovative design in the field of colorectal adenocarcinoma.

### ARTICLE HISTORY

Received 21 December 2022  
Accepted 26 March 2023

### KEYWORDS


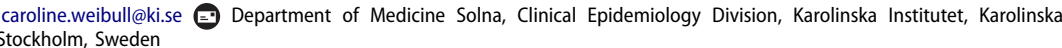
Colorectal cancer; register-based research; database; cancer epidemiology

## Background


Colorectal cancer (CRC) is one of the most common cancer types worldwide. Despite improved treatment strategies including both oncological and surgical therapies, CRC is still one of the leading causes of cancer-related deaths [1]. About 20–25% of patients have metastatic disease at the time of diagnosis and a similar proportion will develop metachronous metastatic disease, which will in turn affect survival negatively [2,3]. Additional research within high quality epidemiological studies is needed to increase the understanding of risk factors, use and outcome of surgical and oncological treatments, and to identify patients that could benefit from more aggressive treatments and more thorough follow-up.

Motivated by a need to improve the quality of treatment for rectal cancer in Sweden, a national quality-of-care

register for rectal cancer was initiated in 1995 [4]. The colon cancer register was added in 2007 [5], together creating the Swedish Colorectal Cancer Register (SCRCR). The SCRCR contains more than 98% of all invasive colorectal adenocarcinomas diagnosed in Sweden [6]. The register includes information on patient and tumor characteristics, diagnostics, treatment, and follow-up, and thus represents an excellent complement to the Swedish Cancer Register (SCR). By linking the SCRCR to not only the SCR, but several other national health-care and sociodemographic registers, a unique database has been created, named CRCBaSe (Colorectal Cancer Database Sweden). CRCBaSe will make it possible to shed light on many unanswered questions regarding CRC in a large unselected cohort of patients, with the aim to improve the future care and outcome of these patients.

**CONTACT** Caroline E. Weibull  [caroline.weibull@ki.se](mailto:caroline.weibull@ki.se) 

\*These authors contributed equally to this work.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/0284186X.2023.2197121>

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

## Material and methods

### *The Swedish Colorectal Cancer Register*

The SCRCR includes rectal cancer diagnoses since 1995 and colon cancer diagnoses from 2007. Only invasive adenocarcinomas, constituting over 95% of all colorectal malignancies, are registered. The register has been described in detail elsewhere [4,5]. The data in SCRCR is registered at the time of diagnosis, and during follow-up (at 30 days, 3, and 5 years after diagnosis, or at relapse if detected) by the treating hospital.

The SCRCR contains information regarding patient and tumor characteristics including tumor location, clinical and pathological stage (TNM 5–10 classifications depending upon year of diagnosis). Information concerning the diagnostic procedures including intention to treatment and assessment of multidisciplinary conference is recorded. Detailed data regarding surgical treatment, radicality and post-operative complications (Clavien-Dindo classification; introduced in the SCRCR in 2011) are noted. Since the register started in 1995, more and more information has been registered, e.g., neoadjuvant and adjuvant treatments are recorded since 2007 and palliative treatments since 2013, making completeness of the register markedly better during more recent years compared to the 1990s and early 2000s. SCRCR has an almost complete coverage, during 2008–2015 SCRCR captured on average 98.5% of colon cancers and 98.8% of rectal cancers registered in the SCR [6]. Several validations of the data in SCRCR have been made [7–9], where the latest showed an average agreement of 90% with patient medical records [6].

For the presentation of CRCBaSe in this study, some variables originating from SCRCR were modified. Disease stage was defined using the pathological classification of T and N status, as well as information on metastatic disease at CRC diagnosis. Moreover, an indicator for *surgically treated with curative intent* was defined based on information on curative intent (original variable in SCRCR) together with if any of the following surgery types had been performed: ileocecal resection, right-sided hemicolectomy, transverse resection, left-sided hemicolectomy, sigmoid resection, total colectomy, anterior resection, abdomino-perineal excision, or Hartmann's procedure. These modifications are also recommended when utilizing CRCBaSe for research purposes.

### *Swedish population-based registers*

The Swedish registers from which data are included in CRCBaSe are held at either the National Board of Health and Welfare (Socialstyrelsen; [www.socialstyrelsen.se](http://www.socialstyrelsen.se)), Statistics Sweden (Statistiska Centralbyrån; [www.scb.se](http://www.scb.se)), or the Swedish Social Insurance Agency (Försäkringskassan; [www.forsakringskassan.se](http://www.forsakringskassan.se)).

#### *Register of Total Population*

The Register of Total Population (RTP) was initiated in 1968 and holds data demographics, such as childbirths, deaths,

civil status, and citizenship. The RTP is the basis for the majority of official statistics on the Swedish population.

#### *Swedish Cancer Register*

The Swedish Cancer Register (SCR) was established in 1958 and contains all primary cancers diagnosed in Sweden. Reporting to the register is done by clinicians, pathologists, and cytologists and is mandatory by law. All cancer diagnoses in the SCR, regardless of year of diagnosis, are (re-)coded according to the 7th revision of the International Classification of Diseases (ICD). Additionally, all tumors are classified according to ICD-9 codes from 1987 and onwards, ICD-10 and ICD-O/2 from 1993, and ICD-O/3 from 2005. The coverage of this register is nearly complete (98%) [10,11].

#### *Cause of Death Register*

Since 1961, the Cause of Death Register (CDR) records data on all deceased Swedish residents who die in Sweden or abroad. Date of death, as well as underlying and contributing causes of death, are recorded.

#### *In- and Outpatient Register*

Six of the Swedish counties (now healthcare regions) started recording inpatient admissions in 1964, initiating the Inpatient Register (IPR). Successively, remaining counties were added and the register reached national coverage in 1987. Since 2001, outpatient visits to non-primary healthcare facilities (including day surgery and psychiatric care), are recorded in the Outpatient Register (OPR).

#### *Prescribed Drug Register*

Since July 2005, all prescribed and dispensed medications are registered in the Prescribed Drug Register (PDR). The register contains information on prescription- and dispense date, ATC code, and dosage. Medications sold without prescription over the counter, or administered at hospitals, are not included in this register.

#### *Medical Birth Register*

Since 1973, all pregnancies resulting in live births in Sweden, and all stillbirths delivered after 28 full gestational weeks (January 1973–June 2008) or 22 gestational weeks (from July 2008), are recorded in the Swedish Medical Birth Register (MBR). The MBR contains information on pregnancy and both maternal and offspring characteristics. For CRCBaSe, data from this register has been added for both mothers (to address questions related to an index person's pregnancies) and offspring (to address question related to an index person's own birth).

#### *Censuses and the LISA database*

Between 1860 and 1990, Sweden conducted population and housing censuses (Folk- och Bostadsräkningarna) to collect information on population size, educational level, and occupation, to name a few. Since 1990, this information is updated annually

and stored within the Longitudinal Integrated Database for Health Insurance and Labour Market Studies (LISA) database.

### Social Security Database (MiDAS)

The Social Security database MiDAS (Mikrodata för analys av Socialförsäkringen) contains information such as sick leave, sickness compensation, and early retirement.

### CRCBaSe record linkages

At birth or immigration to Sweden, all residents are assigned a personal identification number (PIN), unique to that individual [12]. The PIN is used throughout a person's life in all contact with institutions such as healthcare and social services, which in turn enables linkage between population-based registers. For each patient with a record in SCRCR between 1995 (2007 for colon cancer) and 2016, six comparators from the general population were identified. The comparators were matched to the cases (with replacement) on year of birth, sex, county, and being free of CRC at time of patients' diagnosis date (*the index date*). Patients were allowed to act as comparators to another patient at any time point preceding their own CRC diagnosis. For both patients and comparators, first-degree relatives (parents, siblings, and children) were identified and linked.

Finally, information from Swedish population-based registers (described above and illustrated in Figure 1) was merged to each individual in the study population. Based on these linkages, several useful variables are available in CRCBaSe, of

which a selection is described below (and further briefly presented in the Results section).

### Charlson Comorbidity Index

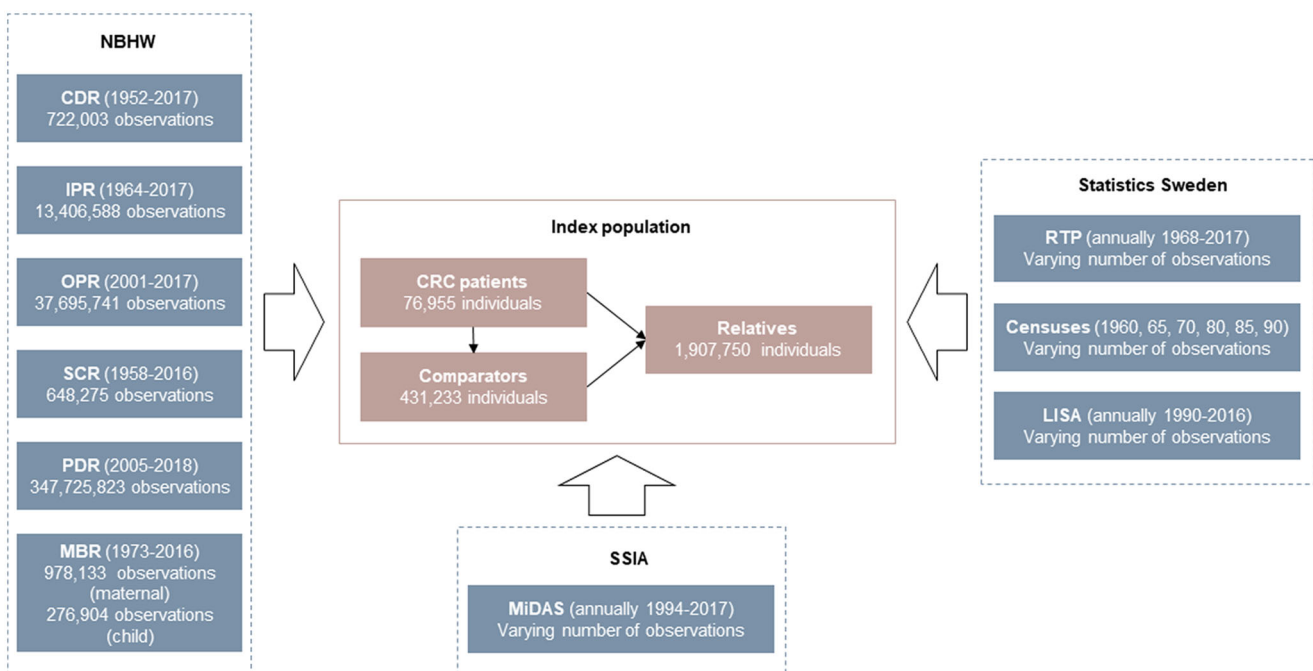
Information on disease history from the SCR and IPR/OPR was used to measure the number of existing comorbidities among both patients and comparators. This was then summarized using the Charlson Comorbidity Index (CCI), calculated according to a recent publication adapted to Swedish registers [13]. For previous cancer diagnoses identified in the SCR (not including any previous CRC), a ten-year period prior to the index date was used, whereas for non-malignant diseases recorded in the IPR/OPR, a five-year period was used for assessment of the CCI.

### Hypertensive disease and diabetes

As examples of specific disease information available in CRCBaSe, occurrences of hypertensive disease and diabetes are identified in the IPR (from ten years to 90 days before index date), the OPR (from five years to 90 days before index date), and the PDR (within 18 months prior to index date). In the present analysis, we have used a broad definition of hypertensive disease based on ICD codes as well as ATC codes, and an occurrence was defined as having at least one diagnosis or one medication dispense.

### Demographical information

Information on country of birth was retrieved from the RTP and categorized into born in Sweden, another Nordic



**Figure 1.** Flowchart of data linkages and sources in CRCBaSe, including the total number of records in each source data. All CRC patients were identified in the Swedish Colorectal Cancer Register (SCRCR), which contains clinical and treatment information. CRCBaSe: Colorectal Cancer Database Sweden; CRC: colorectal cancer; NBHW: National Board of Health and Welfare; CDR: Cause of Death Register; IPR: Inpatient Register; OPR: Outpatient Register; SCR: Swedish Cancer Register; PDR: Prescribed Drug Register; MBR: Medical Birth Register; SSIA: Swedish Social Insurance Agency; MiDAS: Mikro Data for Analysis of the Social Insurance; RTP: Register of Total Population; FoB: Folk- och Bostadsräkningarna (censuses); LISA: Longitudinal Integration Databases for Health Insurance and Labor Market Studies.

country, a non-Nordic country in Europe, and other/unknown country of birth. As measures of socio-economic status, educational level, and/or disposable income can be used. Highest achieved educational level at the time of diagnosis/matching was added from the censuses and LISA database and grouped into  $\leq 9$  years, 10–12 years, and more than 12 years of schooling. Most recent disposable income up until two years prior to diagnosis/matching is also retrieved from the censuses and LISA database, and inflation-adjusted.

The CRCBaSe project has been approved by the Regional Board of the Ethical Committee in Stockholm (DNR: 2014/71-31, 2018/328-32) and by the national Ethical Committee (DNR: 2021-00342).

## Results

Individuals with a rectal cancer diagnosis (1995–2016) or a colon cancer diagnosis (2007–2016) registered in SCRCR were eligible for inclusion in the database and linked to the population-based registers ( $n = 76,955$ ) (Figure 1). Among these, patients were excluded if they did not have a record in the national health-care registers held at the National Board of Health and Welfare ( $n = 59$ ) or if they had a re-used (i.e., non-unique) PIN ( $n = 65$ ). The final study population of CRC patients encompassed 76,831 unique individuals (Supplementary Figure 1). After excluding comparators to whom the CRC patient had been excluded ( $n = 358$ ), a total of 430,875 unique individuals were included as comparators in the database. All comparators had a unique PIN due to the sampling strategy, whereas among eligible parents, children, and siblings, this was not the case. After excluding all relatives with a re-used PIN, a total of 484,859 parents, 931,464 children, and 489,110 siblings were included. To date, CRCBaSe contains 2,413,139 individuals.

A total of 1560 (2.0%) CRC patients had synchronous tumors (defined as having more than one CRC diagnosis

within a time window of six months between the first and last diagnosis). An overview of synchronous tumors occurring in the database is given in Supplementary Table 1. All results from this point forward refer to the first registered tumor in each patient. In case of synchronous CRC tumors diagnosed simultaneously on the first date of diagnosis, colon cancer had priority over rectal cancer.

The annual number rectal cancers varied between approximately 1500 during the early study period and close to 2000 toward the end (Figure 2). For colon cancer, the annual number of cases was around 4000.

The total number of comparator records was 460,976 (Table 1). Because of the later inclusion of colon cancer in the SCRCR, the majority of index dates were 2007 or later. The median age at index date among patients and comparators was 72 and 73 years (inter-quartile range among patients, IQR = 64–80). The proportion of males was 54%. The CCI distribution, and proportions with at least occurrence of hypertensive disease and diabetes were comparable between patients and comparators (Table 1), as was the distribution of birth country and educational level. The vast majority of patients and comparators were born in Sweden (88–89%) and about 20% had more than 12 years of education. The median disposable income was 153,000 SEK/year for both CRC patients and comparators, i.e., lower than in the general population (as these individuals on average are older). Stage registration in the SCRCR was less complete before 2007 (when only rectal cancer patients were included in the SCRCR). However, from 2007 and onwards the stage distribution is as you could expect from a Western population. There were no large differences in stage distribution between sexes, whereas stage III disease was more common among younger patients.

A total of 28,444 (6.6%) comparators were included in more than one stratum (Supplementary Table 2). Five CRC

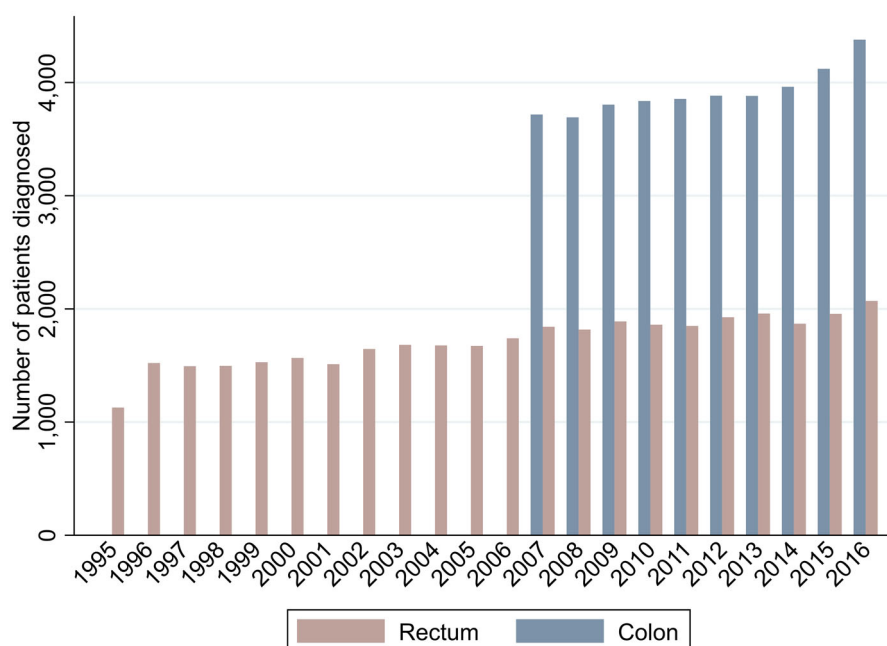


Figure 2. Number of colon- and rectal cancer patients included in the Colorectal Cancer Database Sweden (CRCBaSe) by year of diagnosis.

**Table 1.** Patient and comparator characteristics in the Colorectal Cancer Database Sweden (CRCBaSe), stratified by pathological stage<sup>a</sup> for CRC patients.

	CRC patients					Comparators <i>n</i> (column %)
	Stage I–II <i>n</i> (row %)	Stage III <i>n</i> (row %)	Stage IV <i>n</i> (row %)	Missing <i>n</i> (row %)	All <i>n</i> (column %)	
Overall	28,082 (36.6)	17,094 (22.3)	17,139 (22.3)	14,516 (18.9)	76,831 (100)	460,976 (100)
Calendar year of diagnosis/matching						
1995–2000	860 (9.9)	534 (6.1)	478 (5.5)	6862 (78.6)	8734 (11.4)	52,404 (11.4)
2001–2006	2821 (28.4)	1730 (17.4)	2032 (20.5)	3345 (66.7)	9928 (12.9)	59,566 (12.9)
2007–2012	14,238 (41.9)	8527 (25.1)	9245 (27.2)	1961 (5.8)	33,971 (44.2)	203,826 (44.2)
2013–2016	10,163 (42.0)	6303 (26.1)	5384 (22.3)	2348 (9.7)	24,198 (31.5)	145,180 (31.5)
Age at diagnosis/matching						
≤49	1140 (31.8)	909 (25.3)	997 (27.8)	545 (15.2)	3591 (4.7)	20,244 (4.4)
50–59	2679 (33.5)	1935 (24.2)	1364 (17.1)	1364 (17.1)	7999 (10.4)	45,762 (9.9)
60–69	7100 (36.9)	4553 (23.7)	3045 (15.8)	3045 (15.8)	19,238 (25.0)	112,326 (24.4)
70–79	10,066 (38.7)	5768 (22.2)	4716 (18.1)	4716 (18.1)	26,013 (33.9)	155,754 (33.8)
≥80	7097 (35.5)	3929 (19.7)	4846 (24.2)	4846 (24.2)	19,990 (26.0)	126,890 (27.5)
Median age (IQR)	73 (65–80)	72 (63–79)	71 (63–79)	75 (66–82)	72 (64–80)	73 (65–80)
Sex						
Male	14,898 (36.0)	9008 (21.8)	9407 (22.7)	8058 (19.5)	41,371 (53.9)	248,218 (53.9)
Female	13,184 (37.2)	8086 (22.8)	7732 (21.8)	6458 (18.2)	35,460 (46.2)	212,758 (46.2)
CCI, weighted						
0	17,301 (36.0)	10,976 (22.9)	10,351 (21.6)	9400 (19.6)	48,028 (62.5)	306,445 (66.5)
1	3556 (38.8)	1978 (21.6)	1897 (20.7)	1742 (19.0)	9173 (11.9)	53,669 (11.6)
2+	7225 (36.8)	4140 (21.1)	4891 (24.9)	3374 (17.2)	19,630 (25.6)	100,862 (21.9)
Hypertension						
Yes	21,141 (40.7)	12,178 (23.5)	11,451 (22.1)	7163 (13.8)	51,933 (67.6)	323,310 (70.1)
Diabetes						
Yes	5281 (41.0)	3128 (24.3)	2621 (20.4)	1842 (14.3)	12,872 (16.8)	74,933 (16.3)
Birth country						
Sweden	24,630 (36.2)	14,953 (22.0)	15,114 (22.2)	13,261 (19.5)	67,958 (88.5)	406,320 (88.1)
Other Nordic	1507 (40.2)	829 (22.1)	816 (21.8)	598 (16.0)	3750 (4.9)	21,124 (4.6)
Other European	1373 (37.9)	891 (24.6)	834 (23.1)	521 (14.4)	3619 (4.7)	21,280 (4.6)
Other/unknown	572 (38.0)	421 (28.0)	375 (24.9)	136 (9.0)	1504 (2.0)	12,252 (2.7)
Highest achieved educational level						
≤9 years	11,038 (34.3)	6492 (20.2)	6982 (21.7)	7650 (23.8)	32,162 (41.9)	190,793 (41.4)
10–12 years	10,778 (38.1)	6642 (23.5)	6473 (22.9)	4388 (15.5)	28,281 (36.8)	167,040 (36.2)
>12 years	5908 (39.8)	3738 (25.2)	3427 (23.1)	1758 (11.9)	14,831 (19.3)	92,390 (20.0)
Missing	358 (23.0)	222 (14.3)	257 (16.5)	720 (46.2)	1557 (2.0)	10,753 (2.3)
Disposable income (in units of 1000 SEK)						
Median (IQR)	159 (125–229)	161 (127–235)	156 (124–228)	132 (105–174)	153 (120–219)	153 (120–222)

CRC: colorectal cancer, *n*: number, IQR: inter-quartile range, CCI: Charlson comorbidity index.

Due to rounding, not all percentages add up to 100%.

<sup>a</sup>The stage distribution is based on pTN/ypTN and information on metastatic disease at diagnosis from the SCRCR. Note that as the coverage has improved over the years, the overall proportion with missing stage presented here does not reflect the current quality available in the register.

**Table 2.** Clinical and tumor characteristics for first registered diagnosis among all CRC patients included in the Colorectal Cancer Database Sweden (CRCBaSe), stratified on year of diagnosis for rectal cancer (–2006/2007–).

	Rectum			Total
	Colon	1995–2006	2007–2016	
Overall, <i>n</i> (row %)	39,132 (50.9)	18,662 (24.3)	19,037 (24.8)	76,831 (100)
Disease stage <sup>a</sup> , <i>n</i> (column %)				
Stage I–II	16,505 (42.2)	3681 (19.7)	7896 (41.5)	28,082 (36.6)
Stage III	10,389 (26.6)	2264 (12.1)	4441 (23.3)	17,094 (22.3)
Stage IV	10,031 (25.6)	2510 (13.5)	4598 (24.2)	17,139 (22.3)
Missing	2207 (5.6)	10,207 (54.7)	2102 (11.0)	14,516 (18.9)
ASA score, <i>n</i> (column %)				
Score ≤2	21,232 (54.3)	175 (0.9)	10,828 (56.9)	32,235 (42.0)
Score ≥3	11,234 (28.7)	52 (0.3)	3742 (19.7)	15,028 (19.6)
Not registered <sup>b</sup> /Missing	6666 (17.0)	18,435 (98.8)	4467 (23.5)	29,568 (38.5)
Abdominal resection with curative intent, <i>n</i> (column %)				
Yes	25,541 (65.3)	11,980 (64.2)	12,111 (63.6)	49,632 (64.6)
No/Missing	13,591 (34.7)	6682 (35.8)	6926 (36.4)	27,199 (35.4)

*n*: number, ASA: American Society of Anesthesiologists score.

Due to rounding, not all percentages add up to 100%.

<sup>a</sup>The stage distribution is based on pTN/ypTN and information on metastatic disease at diagnosis from the SCRCR.

<sup>b</sup>ASA score was not registered in the SCRCR 1995–2006, and has thereafter only been registered for patients undergoing major surgery.

patients had less than six comparators identified and matched; one patient had two comparators, and two had four and five comparators, respectively.

Table 2 shows some of the most important clinical and tumor characteristics for the CRC patients according to their first recorded diagnosis, stratified on localization (and for

**Table 3.** Pre- and post-operative treatments and surgical procedures for first registered diagnosis among stage I–III CRC patients in the Colorectal Cancer Database Sweden (CRCBaSe) who were treated with abdominal surgery with curative intent, stratified on localization and year of diagnosis (–2006/2007–) for rectal cancer.

	Colon	Rectum		Total
		1995–2006	2007–2016	
All included patients, <i>n</i> (row %)	23,391 (59.1)	5400 (13.6)	10,813 (27.3)	39,604 (100)
Neoadjuvant treatment, <i>n</i> (column %)				
RT or CRT	118 (0.5)	3438 (63.7)	7113 (65.8)	10,669 (26.9)
CT only	200 (0.9)	10 (0.2)	44 (0.4)	254 (0.6)
None/Missing	23,073 (98.6)	1952 (36.2)	3656 (33.8)	28,681 (72.4)
Surgical procedure, <i>n</i> (column %)				
Ileocecal resection	213 (0.9)	–	–	–
Right-sided hemicolectomy	12,222 (52.3)	–	–	–
Transverse resection	377 (1.6)	–	–	–
Left-sided hemicolectomy	2635 (11.3)	–	–	–
Sigmoid resection	5234 (22.6)	–	–	–
Total colectomy	991 (4.2)	–	–	–
Anterior resection	1178 (5.0)	3322 (61.5)	5733 (53.0)	10,233 (25.8)
Abdomino-perineal excision	29 (0.1)	1503 (27.8)	3847 (35.6)	5379 (13.6)
Hartmann's procedure	512 (2.2)	575 (10.7)	1189 (11.0)	2276 (5.8)
Pathological T stage, <i>n</i> (column %)				
T0	–	–	–	20 (0.1)
T1	1616 (6.7)	400 (7.4)	927 (8.6)	2943 (7.4)
T2	3520 (14.5)	1589 (29.4)	3287 (30.4)	8396 (21.2)
T3	14,452 (60.9)	3224 (59.7)	5931 (54.9)	23,607 (59.6)
T4	3799 (17.5)	186 (3.4)	651 (6.0)	4636 (11.7)
TX	–	–	–	–
Missing	–	–	–	–
Pathological N stage, <i>n</i> (column %)				
N0	14,610 (62.5)	3416 (63.3)	6761 (62.5)	24,787 (62.6)
N1	5735 (24.5)	1250 (23.2)	2697 (24.9)	9682 (24.5)
N2	3027 (12.9)	733 (13.6)	1354 (12.5)	5114 (12.9)
NX	–	–	–	20 (0.1)
Referred to adjuvant chemotherapy, <i>n</i> (column %)	6098 (23.9)	20 (0.2)	2523 (20.8)	8641 (17.4)
Local recurrence, <i>n</i> (column %)	923 (3.6)	772 (6.4)	448 (3.7)	2143 (4.3)
Distant metastatic disease, <i>n</i> (column %)	4028 (15.8)	2497 (20.8)	2486 (20.5)	9011 (18.2)
Follow-up time in years, median (IQR)	10.1 (7.6–12.7)	18.3 (16.8–21.1)	10.2 (7.8–12.8)	10.9 (8.1–14.0)
Median survival time in years (95% CI)	10.3 (10.2–10.6)	8.9 (8.5–9.2)	12.1 (11.8–12.5)	10.7 (10.5–10.8)

CRC: colorectal cancer, *n*: number, RT: radiotherapy, CT: chemotherapy, CRT: both chemo- and radiotherapy or concomitant chemoradiotherapy, IQR: inter-quartile range, CI: confidence interval.

Due to rounding, not all percentages add up to 100%. Cell counts less than five are not shown.

rectal cancer also on calendar period of diagnosis: 1995–2006/2007–2016). In 2007, colon cancer was added to the register, the number of registered variables increased substantially, and the name of the register was changed to SCRCR. The distribution of colon and rectal cancer across all years was similar, with 39,132 (51%) first colon and 37,699 (49%) first rectal cancers, whereas during the years where both tumor locations were included (2007–2016), the majority was colon cancers (67%). Among colon cancer patients, the stage distribution was 42% stage I–II, 27% stage III, and 26% stage IV. For rectal cancer patients diagnosed 2007–2016, stage was distributed as 42%, 23%, and 24% for stage I–II, III, and IV, respectively.

The number of stage I–III colon cancer patients who were treated with abdominal surgery with curative intent was 23,391 (60% out of total number of colon cancers) (Table 3). For rectal cancer patients among whom this subset could be constructed, the corresponding numbers were 5400 (29%) in 1995–2006 and 10,813 (57%) in 2007–2016. In these groups of patients, the most common surgical procedures were right-sided hemicolectomy for colon cancer (52%) and anterior resection for rectal cancer (62% in 1995–2006 and 53% in 2007–2016). Very few of the curatively surgically treated stage I–III colon cancer patients received neoadjuvant treatment, whereas among rectal cancer patients diagnosed 2007

or later, 66% received this treatment. The proportion among these patients who were referred to adjuvant chemotherapy was 24% in colon and 21% in rectal cancer patients diagnosed 2007 or later. Colon cancer patients and rectal cancer patients diagnosed from 2007 had similar occurrence of registered local recurrences (around 3.6%) and distant metastatic disease (colon 16% and rectal 2007–2016, 21%). Follow-up for death and relapse was available until 31st of May 2022, resulting in a median follow-up of 10.9 years (IQR = 8.1–14.0). The median survival time (the time point at which 50% of patients were still alive), was 10.3 years (95% CI: 10.2–10.6) for colon cancer patients and 12.1 years (95% CI: 11.8–12.5) for rectal cancer patients diagnosed during the latter period.

Overall survival, estimated using the Kaplan–Meier method, among comparators and patients (stratified by stage) is presented in Supplementary Figure 2.

## Discussion

CRCBaSe is a large register linkage containing almost 80,000 patients with CRC, population-based comparators, and relatives to both patients and comparators. This unique database with more than 2.4 million individuals is expected to become

an important asset in CRC research in the future. There are several ongoing studies in CRCBaSe, and a few recently published [14–18].

Strengths with CRCBaSe include its national coverage with high quality data on patient and cancer characteristics including treatment and cancer recurrence, information that is not available in the SCR. Linkage to the other national registers available enables not only adjustments for potential confounders (such as socioeconomic status, comorbid conditions, and demographics), but also increases the amount of information for each individual considerably and assures more complete follow-up. The steering committee for CRCBaSe consists of colorectal surgeons, clinical oncologists, epidemiologists, and a biostatistician, to ensure competence and expertise in CRC and treatment thereof, in register-based research, and statistical methods.

CRCBaSe is limited by only including colon cancer cases from 2007, and also by outpatient registrations (which includes specialist care) starting in 2001 and the registration of drug dispenses starting in 2005. Still, as a large number of patients were diagnosed during more recent years when coverage in all included registries is more complete, we believe this issue will have limited impact on future studies. While the follow-up in the SCRCR has recently been updated to include relapses and deaths until 31st of May 2022, the database currently only consists of patients diagnosed 2016 or earlier, and the other linked registers include information until 2016–2018 (depending on register). Regular updates of the existing linkages are therefore planned, and the first one underway is expected to be finalized during spring 2023. It should also be noted that the number of variables registered in the SCRCR, thus possible to study, has continuously increased over time why some research questions are not possible to address for patients registered early on. E.g., details on oncological treatments besides surgery are limited prior to 2014. Data of molecular properties is not available prior to 2017. Information on whether (neo-)adjuvant treatment was given is available from 2007, but also after this time point limited to surgeon registration of whether referral to such treatment was done or not. Depending upon the year of diagnosis, this information is variable and, thus, uncertain.

The occurrence of synchronous tumors in CRCBaSe was 2%, slightly lower than previous studies that have shown about 3% [19,20], but in line with the prevalence range in Europe (1.1–8.1%) reported in a review from 2014 [21].

The proportion of registered recurrences in CRCBaSe is similar to what has been reported previously. A recent publication showed a high (from an international perspective) proportion of recurrences registered in the SCRCR during 2010–2018 [22]. As registration was more accurate with longer follow-up (due to the fixed time points of follow-up at three and five years) revealing a delay in reporting that might not be captured in the present version. The information on recurrences will most likely be improved in updated versions of CRCBaSe.

CRCBaSe offers great advantages in comparison to utilizing the SCRCR in isolation (i.e., without linkages to other

register data), which increases research questions that can be studied markedly and improves quality as well. Surgery codes and dates in the IPR can be used to complement and correct information on CRC surgery as registered in the SCRCR. For example, only one surgical procedure is recorded in SCRCR. By using information from the IPR, additional surgical codes can be retrieved, yielding more accurate information on performed surgery. Additionally, patients who have a CRC diagnosis prior to 1995 (or 2007 for colon cancer) which is not captured by the SCRCR can via the database easily be handled (excluded or noted) using their information in the SCR.

In conclusion, CRCBaSe is a unique resource for real-world CRC research. This national, high-quality database research platform will make it possible to address clinically relevant research questions that have not been possible to fully explore previously.

## Acknowledgements

This project was made possible by the continuous work of steering groups of the Swedish Colorectal Cancer Register (SCRCR) and Colorectal Cancer Database Sweden (CRCBaSe). We thank Björn Enhus at the National Board of Health and Welfare, Jonas Färnstrand at Statistics Sweden, and Claes Edwardsson at the Swedish Social Insurance Agency.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

The CRCBaSe project was supported financially by the Swedish Cancer Society, the Stockholm Cancer Society, and the Regional Agreement on Medical Training and Clinical Research between the Stockholm County Council and Karolinska Institutet (ALF-project).

## ORCID

Caroline E. Weibull  <http://orcid.org/0000-0001-8231-8183>  
 Bengt Glimelius  <http://orcid.org/0000-0002-5440-791X>  
 Ingvar Syk  <http://orcid.org/0000-0002-9314-6919>  
 Peter Matthiessen  <http://orcid.org/0000-0003-4939-4189>  
 Caroline Nordenvall  <http://orcid.org/0000-0002-5112-8894>  
 Anna Martling  <http://orcid.org/0000-0001-5998-4735>

## Data availability statement

The data underlying this study are available on request from the steering group of SCRCR and CRCBaSe with the appropriate approvals, but restrictions apply. Information about the application process is available on CRCBaSe's webpage: <https://cancercentrum.se/samverkan/cancerdiagnoser/tjocktarm-andtarm-och-anal/tjock-och-andtarm/kvalitetsregister/forskning/forskningsdatabas/>

## References

- [1] Bray F, Ferlay J, Soerjomataram I, et al. Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424.

- [2] Benitez Majano S, Di Girolamo C, Rachet B, et al. Surgical treatment and survival from colorectal cancer in Denmark, England, Norway, and Sweden. *Lancet Oncol.* 2019;20(1):74–87.
- [3] Osterman E, Glimelius B. Recurrence risk after up-to-date colon cancer staging, surgery, and pathology: analysis of the entire Swedish population. *Dis Colon Rectum.* 2018;61(9):1016–1025.
- [4] Pahlman L, Bohe M, Cedermark B, et al. The Swedish Rectal Cancer Registry. *Br J Surg.* 2007;94(10):1285–1292.
- [5] Kodeda K, Nathanaelsson L, Jung B, et al. Population-based data from the Swedish Colon Cancer Registry. *Br J Surg.* 2013;100(8):1100–1107.
- [6] Moberger P, Skoldberg F, Birgisson H. Evaluation of the Swedish Colorectal Cancer Registry: an overview of completeness, timeliness, comparability and validity. *Acta Oncol.* 2018;57(12):1611–1621.
- [7] Jorgren F, Johansson R, Damber L, et al. Risk factors of rectal cancer local recurrence: population-based survey and validation of the Swedish Rectal Cancer Registry. *Colorectal Dis.* 2010;12(10):977–986.
- [8] Jorgren F, Johansson R, Damber L, et al. Validity of the Swedish Rectal Cancer Registry for patients treated with major abdominal surgery between 1995 and 1997. *Acta Oncol.* 2013;52(8):1707–1714.
- [9] Gunnarsson U, Seligsohn E, Jestin P, et al. Registration and validity of surgical complications in colorectal cancer surgery. *Br J Surg.* 2003;90(4):454–459.
- [10] Barlow L, Westergren K, Holmberg L, et al. The completeness of the Swedish Cancer Register: a sample survey for year 1998. *Acta Oncol.* 2009;48(1):27–33.
- [11] Mattsson B, Rutqvist LE, Wallgren A. Undernotification of diagnosed cancer cases to the Stockholm Cancer Registry. *Int J Epidemiol.* 1985;14(1):64–69.
- [12] Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, et al. The Swedish Personal Identity Number: possibilities and pitfalls in health-care and medical research. *Eur J Epidemiol.* 2009;24(11):659–667.
- [13] Ludvigsson J, Appelros P, Askling J, et al. Adaptation of the Charlson Comorbidity Index for register-based research in Sweden. *Clin Epidemiol.* 2021;13:21–41.
- [14] Shahrivar M, Weibull CE, Ekstrom Smedby K, et al. Low-dose aspirin use and colorectal cancer survival in 32,195 patients—a national cohort study. *Cancer Med.* 2022;12(1):315–324.
- [15] Ljunggren M, Weibull CE, Rosander E, et al. Hospital factors and metastatic surgery in colorectal cancer patients: a population-based cohort study. *BMC Cancer.* 2022;22(1):907.
- [16] Ljunggren M, Weibull CE, Palmer G, et al. Sex differences in metastatic surgery following diagnosis of synchronous metastatic colorectal cancer. *Int J Cancer.* 2022;152(3):363–373.
- [17] Lundqvist E, Myrberg IH, Boman SE, et al. Autoimmune and metabolic diseases and the risk of early-onset colorectal cancer: a nationwide nested case-control study. *Cancers.* 2023;15(3):688.
- [18] Rutegård M, Moshtaghi-Svensson J, Weibull CE, et al. Exposure to oestrogen and risk of anastomotic leakage after colorectal cancer surgery – a clue to the different leak rates in men and women. *Colorectal Dis.* 2023;25(1):9–15.
- [19] Lee BC, Yu CS, Kim J, et al. Clinicopathological features and surgical options for synchronous colorectal cancer. *Medicine.* 2017;96(9):e6224.
- [20] Welch JP. Multiple colorectal tumors. An appraisal of natural history and therapeutic options. *Am J Surg.* 1981;142(2):274–280.
- [21] Lam AK, Chan SS, Leung M. Synchronous colorectal cancer: clinical, pathological and molecular implications. *World J Gastroenterol.* 2014;20(22):6815–6820.
- [22] Osterman E, Hammarstrom K, Imam I, et al. Completeness and accuracy of the registration of recurrences in the Swedish Colorectal Cancer Registry (SCRCR) and an update of recurrence risk in colon cancer. *Acta Oncol.* 2021;60(7):842–849.