

Incidence, prevalence, and survival of colorectal cancer in the United Kingdom from 2000-2021: a population-based cohort study

Authors:

Patricia Pedregal-Pascual*^{1,2}

Carlos Guarner-Argente*¹

Eng Hooi Tan³

Asieh Golozar^{4,5}

Talita Duarte-Salles^{6,7}

Andreas Weinberger Rosen⁸

Antonella Delmestri³

Wai Yi Man³

Edward Burn³

Daniel Prieto Alhambra^{3,7}

Danielle Newby³

* These authors contributed equally to this work

Affiliations:

¹ Gastroenterology Department, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain.

² Medicine Department, Universitat Autònoma de Barcelona, Barcelona, Spain

³ Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

⁴ Odysseus Data Service, Cambridge, MA USA

⁵ OHDSI Center at the Roux Institute, Northeastern University, Boston, MA USA

⁶ Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain

⁷ Department of Medical Informatics, Erasmus University Medical Centre, Rotterdam, The Netherlands

⁸ Center for Surgical Science, Zealand University Hospital, Lykkebækvej 1, 4600, Køge, Denmark

Corresponding author(s):

Professor Daniel Prieto Alhambra

Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK.

daniel.prietoalhambra@ndorms.ox.ac.uk

ORCID: <https://orcid.org/0000-0002-3950-6346>

KEYWORDS:

colorectal cancer, gastrointestinal cancer, oncology, incidence, prevalence, cancer survival, population-based study

ABSTRACT

Background: The management of colorectal cancer (CRC) is evolving, with advances in screening and treatment. We leveraged population-based data to generate up-to-date UK estimates of age and sex-specific incidence and prevalence and overall survival for the period 2000-2021.

Methods: We analysed nationally representative primary care records from Clinical Practice Research Datalink (CPRD) GOLD, and replicated in CPRD Aurum. We calculated incidence rates, prevalence, and short- and long-term survival stratified by age, sex, and diagnosis year.

Results: Unadjusted prevalence increased in the study period, from 15.6 to 46.4/10,000. Overall incidence was 61.5/100,000 person years, increasing in 2000-2011 to drop slightly in 2011-2014, and then plateauing. In contrast, early-onset CRC raised uninterruptedly throughout the study period, from 8.33 to 19.07/100,000 person-years.

Overall survival was 78.3%, 51.4% and 38.5% at 1-, 5-, and 10-years respectively, lower in men compared to women. Modest improvements in survival were observed over the study period, particularly for 60-69 year old patients.

Conclusion: The overall prevalence of CRC in the UK has tripled in the last 20 years, leading to increased healthcare resource needs and with slight survival improvements. A worrying increasing trend of early-onset CRC is observed, warranting further research into its diagnosis and management.

ABBREVIATIONS:

CPRD, Clinical Practice Research Datalink

CRC, colorectal cancer

EOCRC, early-onset colorectal cancer

FIT, Faecal Immunochemical Test

gFOBT, guaiac Faecal Occult Blood Test

IR, incidence rate

IQR, interquartile range

KM, Kaplan-Meier

NHS, National Health Service

OS, overall survival

PP, period prevalence

Pys, person-years

SNOMED CT, Systematized Nomenclature of Medicine - Clinical Terms

INTRODUCTION

Colorectal cancers (CRC) are the third most common cancer (more than 1.9 million cases) and the second most common cause of cancer-related death (935,000 deaths) for both men and women worldwide in 2020 ¹. From 2016–2018 there were around 42,900 new CRC cases in the United Kingdom (UK) every year, being the fourth most common cancer in this country ².

The majority of CRC are sporadic and follow the adenoma-carcinoma sequence over a period of 10-15 years ³. There are several well-characterized risk factors such as age, family history and hereditary cancer syndromes, inflammatory bowel diseases, metabolic syndrome, sedentary lifestyle, diet, heavy alcohol consumption and smoking ³.

The risk of developing CRC increases with advancing age where 90% of global cases and deaths occur from age 50 years ³, although there has been an increase in the number of diagnoses in those under 50 years of age ⁴. CRC is more common in males ³ and reasons for these sex differences are potentially due to males having a higher vulnerability to environmental risk factors and higher exposure to them as well as lower uptake of CRC screening based on faecal samples ³. Additionally, males do not benefit from the potential protective effect of endogenous oestrogen, which has been linked to a reduction in risk in females ^{5,6}.

Considering the risk factors above, there are opportunities for intervention to prevent CRC onset (primary prevention), achieve early detection (secondary prevention), or mitigate its impact on prognosis (tertiary prevention), thereby rendering it a potentially preventable disease ⁷. CRC screening has been demonstrated to be an effective and cost-effective approach ³ which has been implemented by numerous public health organizations worldwide including the UK ⁸. These screening programmes aim to reduce the incidence of CRC by diagnosing colorectal adenomas before CRC onset and improving prognosis by earlier detection ⁷. In the UK, the NHS (National Health Service) Bowel Cancer Screening Programme began in 2006 ⁸ and is offered every 2 years to those aged 60 to 74 with the screening now being offered from 50 years of age ⁹. The effectiveness of CRC screening has been reflected in improvements in overall survival where in England, the 5-year survival in those diagnosed through the screening programme is 81.9% but drops to 53.0% in non-screen-detected cases ¹⁰.

Other population-based studies using cancer registries have shown that overall incidence in UK has increased but the trends indicate IRs have declined in recent years, especially after 2010, with similar trends in both sexes ^{11,12}. These studies also showed an increase in the incidence of CRC in people younger than 50 years and older than 75 years of age ¹¹⁻¹³. Furthermore, evidence suggests that CRC mortality in the UK has declined between 2000 to 2017 ¹³. However, despite this, the UK has one of the poorest survival rates compared to the rest of Europe and other high-income countries worldwide ^{14,15}.

Understanding trends in the incidence, prevalence, and survival of CRC is an important aspect to assess the efficacy of cancer diagnosis and management to inform the development of effective preventive and screening strategies. Due to this and changes in screening age, tobacco and alcohol usage and increases of CRC in younger adults in recent years, a comprehensive assessment of the trends of CRC in relation to different population strata is lacking in the UK. Therefore, the aim of this study is to describe the population-based burden of CRC and secular trends in terms of incidence, prevalence, and survival from 2000-2021 using two primary care databases from the UK.

METHODS

Study design, setting, and data sources

We carried out a population cohort study using routinely collected primary care data from the UK. People with a diagnosis of CRC and a background cohort (denominator population) were identified from Clinical Practice Research Datalink (CPRD) GOLD database (July 2022) to estimate overall survival, incidence, and prevalence. We additionally carried out this study using CPRD Aurum (June 2021) to compare the results obtained from GOLD. Both these databases contain pseudonymised patient-level information on demographics, lifestyle data, clinical diagnoses, prescriptions, and preventive care provided to patients and collected by the NHS (National Health Service) as part of their care and support. CPRD GOLD contains data from across the UK whereas Aurum only contains data from England. Both databases are established primary care databases covering together over 50 million people and both were mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) ¹⁶.

Study participants and time at risk

For the incidence and prevalence analysis, the study cohort consisted of individuals present in the database from 1st January 2000. All participants were required to be 18 years or older and have at least one year of prior history. For CPRD GOLD, these individuals were followed up to whichever came first: the cancer outcome of interest, exit from the database, date of death, or the 31st of December 2021 (the end of study period) whereas for Aurum, the end of the study period was 31st of December 2019. We studied until 2019 in CPRD Aurum as there were problems with data quality of this database post 2019 that were later resolved after this study was completed.

For the survival analysis, only individuals with a newly diagnosed cancer were included. These individuals were followed up from the date of their diagnosis to either date of death, practice stopped contributing to the database, participant left the practice or end of the study period. Any patients whose death and cancer diagnosis occurred on the same date were excluded from the survival analysis.

Colorectal cancer definition

We used Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) diagnostic codes to identify primary CRC events. Specific diagnostics codes for malignancies of the colon and/or the rectum were included as well as unspecific codes covering these locations as these were assumed to be most likely to be from the colon or rectum. Diagnostic codes indicative of non-malignant cancer, melanoma and lymphoma were excluded as well as codes related to cancers of the appendix. For incidence and survival analyses, only diagnostic codes defining incidence disease were included whereas for prevalence estimates diagnostics codes indicative of recurrence or metastatic status of the disease were also included. The CRC cancer definition was reviewed with the aid of the CohortDiagnostics R package ¹⁷. This package was used to identify additional codes of interest and to remove those highlighted as irrelevant based on feedback from clinicians with oncology, primary care, and real-world data expertise through an iterative process during the initial stages of analyses. The clinical code lists used to define CRC can be found in the supplementary information S1. For survival analysis, mortality was defined as all-cause mortality based on records of date of death. OMOP-based computable phenotypes are available, together with all analytical code in

GitHub to enable research reproducibility (<https://github.com/oxford-pharmacoepi/EHDENcancerIncidencePrevalence>)

Statistical methods

The population characteristics of patients with a diagnosis of CRC were summarised, with median and interquartile range (IQR) used for continuous variables, while counts and percentages were used for categorical variables.

We calculated the crude overall and annualised incidence rates (IR) and period prevalence (PP) for CRC from 2000 to 2021 in GOLD. For incidence, the number of events, the observed time at risk, and the incidence rate per 100,000 person years were summarised along with 95% confidence intervals (95% CI). Annualised IRs were calculated as the number of incident CRC cases as the numerator and the actual and exact recorded number of person-years (pys) in the general population within that year as the denominator whereas overall incidence was calculated from 2000 to 2021.

Age-standardized IRs were calculated using the 2013 European Standard Population (ESP2013)¹⁸. The ESP2013 serves as a population standard with a predefined age distribution to account for differences in age structures between different populations. The ESP2013 provides predefined age distribution in five-year age bands; therefore, we collapsed these to 10-year age bands. We used the age distribution of 20-29 years from ESP2013 for age-standardization as age distributions were not available for the 18-29 years age band used in this study.

Period prevalence was calculated on 1st January for the years 2000 to 2021, with the number of patients fulfilling the case definition for CRC as the numerator. Anyone with a diagnosis of CRC was included and followed until the end of their observation period. We estimated total prevalence as although many of these cases may be considered cured after five/ten years and no longer being actively treated, people in this survivorship phase may have long-term medical needs, and therefore we considered it is important to provide accurate counts to allow for healthcare planning. The denominator was the number of participants on the 1st of January in the respective years for each database. The number of events, and prevalence (%) were summarised along with 95% confidence intervals.

For survival analysis, we used the Kaplan-Meier (KM) method to estimate the overall survival probability from observed survival times with 95% confidence intervals. We estimated the median survival and survival probability 1, 5, and 10 years after diagnosis. All results were stratified by database, by age (10-year age bands apart from the first and last age bands which were 18 to 29 years and 90 years and older respectively) and sex. Additionally, for CPRD GOLD, we stratified by calendar time of cancer diagnosis (2000-2004, 2005-2009, 2010-2014, 2015-2019 and 2020-2021). To avoid possible patient identification, we do not report results with fewer than five cases.

For Aurum, the same statistical analyses were performed using data from 1st January 2000 to 31st December 2019 to compare the results obtained from GOLD.

The statistical software R version 4.2.3 was used for analyses. For calculating incidence and prevalence, we used the IncidencePrevalence R package¹⁹. For survival analysis we used the survival R package²⁰.

RESULTS

Patient Populations and characteristics

Overall, there were 24,340,860 and 11,388,117 eligible people available from CPRD Aurum and GOLD respectively. Attrition tables can be found in the supplementary information (Supplement S2). A summary of patient characteristics of those with a diagnosis of CRC for both databases is shown in Table 1.

Table 1: Baseline characteristics of colorectal cancer patients at the time of diagnosis for each database.

Database	CPRD Aurum	CPRD GOLD
Number of CRC patients	99321	53797
Sex: Male (N[%])	55510 (55.9%)	29800 (55.4%)
Age (years) (Median [IQR])	72 (63 to 80)	72 (63 to 80)
Age Groups N (years) (%)		
18-29	231 (0.2%)	123 (0.2%)
30-39	1182 (1.2%)	524 (1.0%)
40-49	3727 (3.8%)	2001 (3.7%)
50-59	11432 (11.5%)	6367 (11.8%)
60-69	24134 (24.3%)	13466 (25.0%)
70-79	31736 (32.0%)	17286 (32.1%)
80-89	22928 (23.1%)	12118 (22.5%)
90+	3951 (4.0%)	1912 (3.6%)
Prior history, days		
median [IQR]	6623 (3,363 to 11,038)	3493 (1,892 to 5,250)
General conditions (any time prior)		
Atrial fibrillation	8023 (8.1%)	3747 (7.0%)
Cerebrovascular disease	6730 (6.8%)	3288 (6.1%)
Chronic liver disease	417 (0.4%)	201 (0.4%)
Chronic obstructive lung disease	6775 (6.8%)	3365 (6.3%)
Coronary arteriosclerosis	1133 (1.1%)	547 (1.0%)
Dementia	1526 (1.5%)	774 (1.4%)
Depressive disorder	10547 (10.6%)	5352 (9.9%)
Diabetes mellitus	14553 (14.7%)	6489 (12.1%)
Gastrointestinal haemorrhage	19013 (19.1%)	10245 (19.0%)
Heart failure	3672 (3.7%)	1796 (3.3%)
Hyperlipidemia	10411 (10.5%)	5009 (9.3%)
Hypertensive disorder	40764 (41.0%)	14932 (27.8%)
Ischemic heart disease	12417 (12.5%)	5495 (10.2%)
Osteoarthritis	25022 (25.2%)	10583 (19.7%)
Peripheral vascular disease	2051 (2.1%)	1017 (1.9%)
Pulmonary embolism	1699 (1.7%)	724 (1.3%)
Renal impairment	13117 (13.2%)	6891 (12.8%)
Venous thrombosis	5749 (5.8%)	2848 (5.3%)

IQR: interquartile range

Overall, those with CRC are more likely to be male (~56%), with a median age of 72 years old across both databases. The highest percentage of patients were those aged 70-79 years old, contributing to 32% of diagnosed patients.

Incidence rates stratified by calendar year, age and sex.

The overall crude IR of CRC from 2000 to 2021 was 61.5 (95% CI 61.0–62.0) per 100000 person-years for GOLD with a slightly lower IR for Aurum (56.6 95%CI 56.2 to 56.9). For GOLD, sex-specific IRs for males and females were 69.3 (95% CI 68.5–70.1) and 54.0 (95% CI 53.3–54.7) per 100000 person-years respectively, with similar results obtained for Aurum.

Annualised IRs increased across the study period for the whole population and both sexes with males having higher IRs than females (Figure 1). For males, IRs declined from 2011 to 2014-5 but increased up to 2019 for both databases. Whereas for females, in GOLD, IRs peaked in 2011 before dropping until 2019, while in Aurum IRs remained stable with a gradual increase towards the end of the study period. For GOLD there was a drop in IR in 2020 but an increase in IR in 2021 across the whole population and sex. Age standardized results using the European Standard Population for 2013 for GOLD for both sexes can be found in the supplement (Supplement S3). All results for this study can be found and downloaded in a user-friendly interactive web application: <https://dpa-pde-oxford.shinyapps.io/ColorectalCancerIncPrevSurvShiny/>.

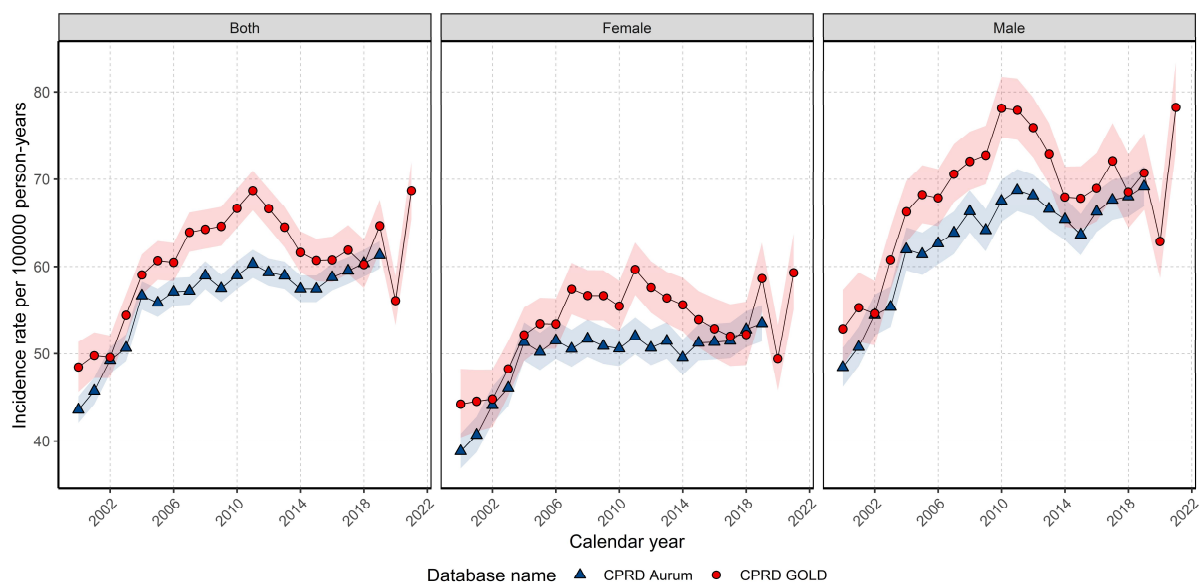


Figure 1: Crude annualised incidence rates for CRC from 2000 to 2021 (GOLD) and 2000-2019 (Aurum) stratified by database and sex.

Overall IRs increased with age up to 80–89 years. Those aged 18 to 29 had the lowest overall IRs across the study period with IRs of 0.77 (95% CI 0.64 to 0.92) per 100,000 person years in GOLD, whereas those aged 80–89 years-old had the highest with IRs of 277.2 (95% CI 272.3-282.1). Similar results were obtained in Aurum (Supplement S4).

Annualised IRs for each age group (Figure 2) show IRs have increased annually in those aged 30 to 59 across both databases. For those aged 60-79 years old, IRs rose and peaked in 2011 before falling and remaining stable until the end of the study period. IRs for those aged 80-89 years old remained stable after 2006. For those aged 90 and older IRs in GOLD did not increase from 2011 to the end of the study period, however in Aurum, IRs increased gradually over the study period. Across most age groups in GOLD, IRs in 2020 decreased before increasing in 2021. Stratification on both sex and age group showed similar trends in Figure 2

for both sexes. However, males had higher incidence across the study period apart from those younger than 50 years of age where there were no sex differences (Supplement S5).

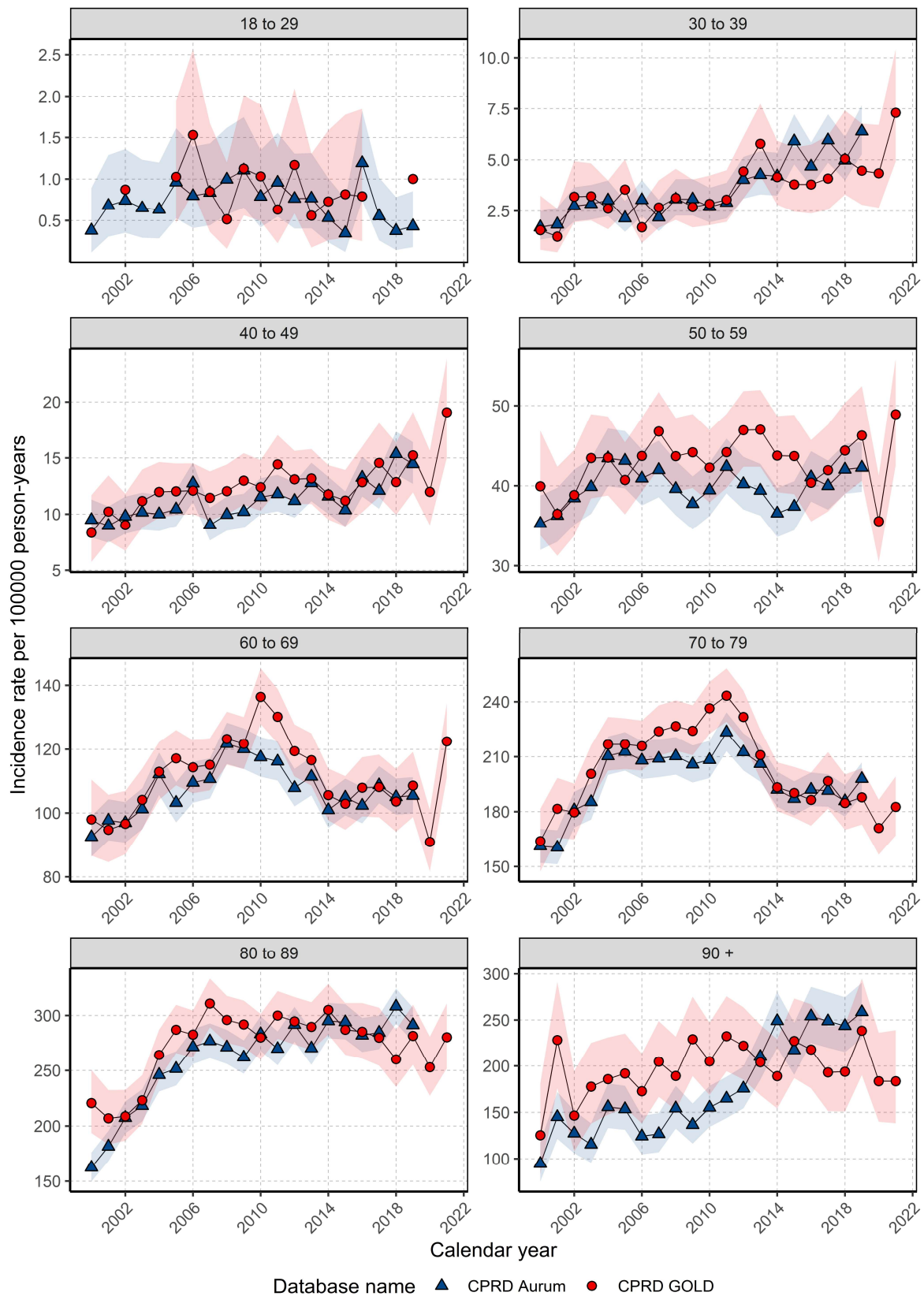


Figure 2: Annualised incidence rates from 2000 to 2021 stratified by database and age group.

Prevalence for study population with database, age, and sex stratifications

In GOLD, PP for CRC in 2021 was 0.46% (0.46% to 0.47%) with similar PP in 2019 when comparing GOLD and Aurum (0.44% GOLD, 0.42% Aurum). PP in 2021 was 1.23-fold higher for males compared to females in GOLD (Males 0.51%, Females 0.42%). Annualised PP increased over the study period (Figure 3) with PP increasing 1.91-fold (Aurum) and 2.97-fold (GOLD) for the whole population, with similar increases for both females and males (Figure 3).

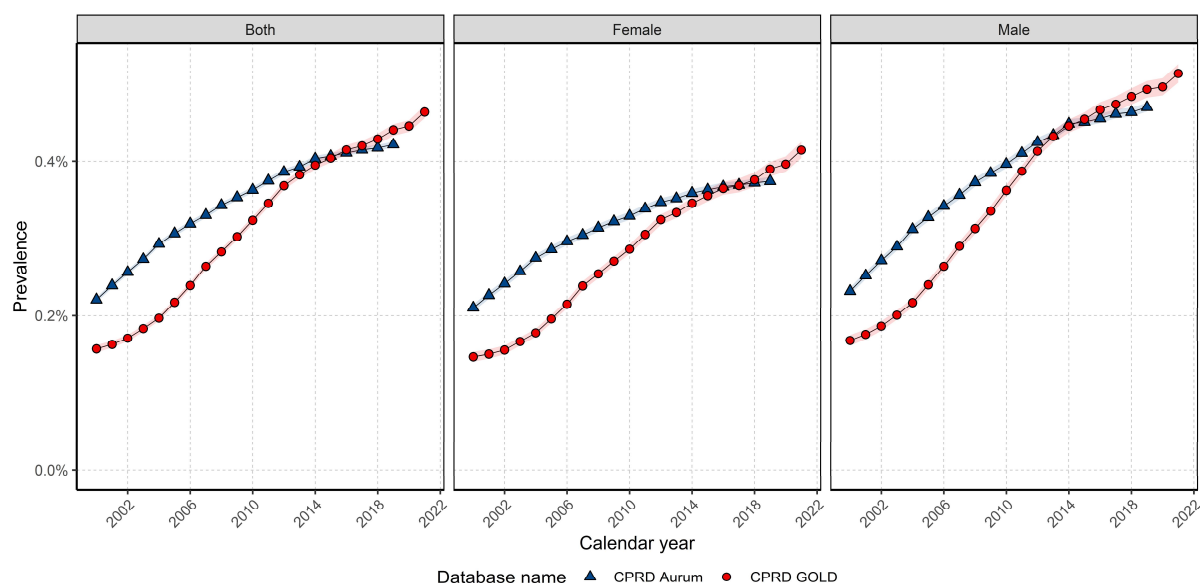


Figure 3: Annualised prevalence from 2000 to 2021 for whole population and stratified by sex.

When stratifying by age group, PP was highest with increasing age and was highest in those aged 80 years and older (Figure 4). PP increased then plateaued from 2011-12 for those aged between 60 to 79 years old whereas PP increased until 2007 with a small gradual decline towards the end of the study period in those aged 18 to 29 years old. Those aged 30-49 years of age and over 80 years of age showed an increase in PP across the whole study period across both databases. Stratification on both sex and age group showed similar trends to Figure 4 however males had higher prevalence across the study period compared to females apart from those younger than 50 years of age where there were no sex differences (Supplement S6).

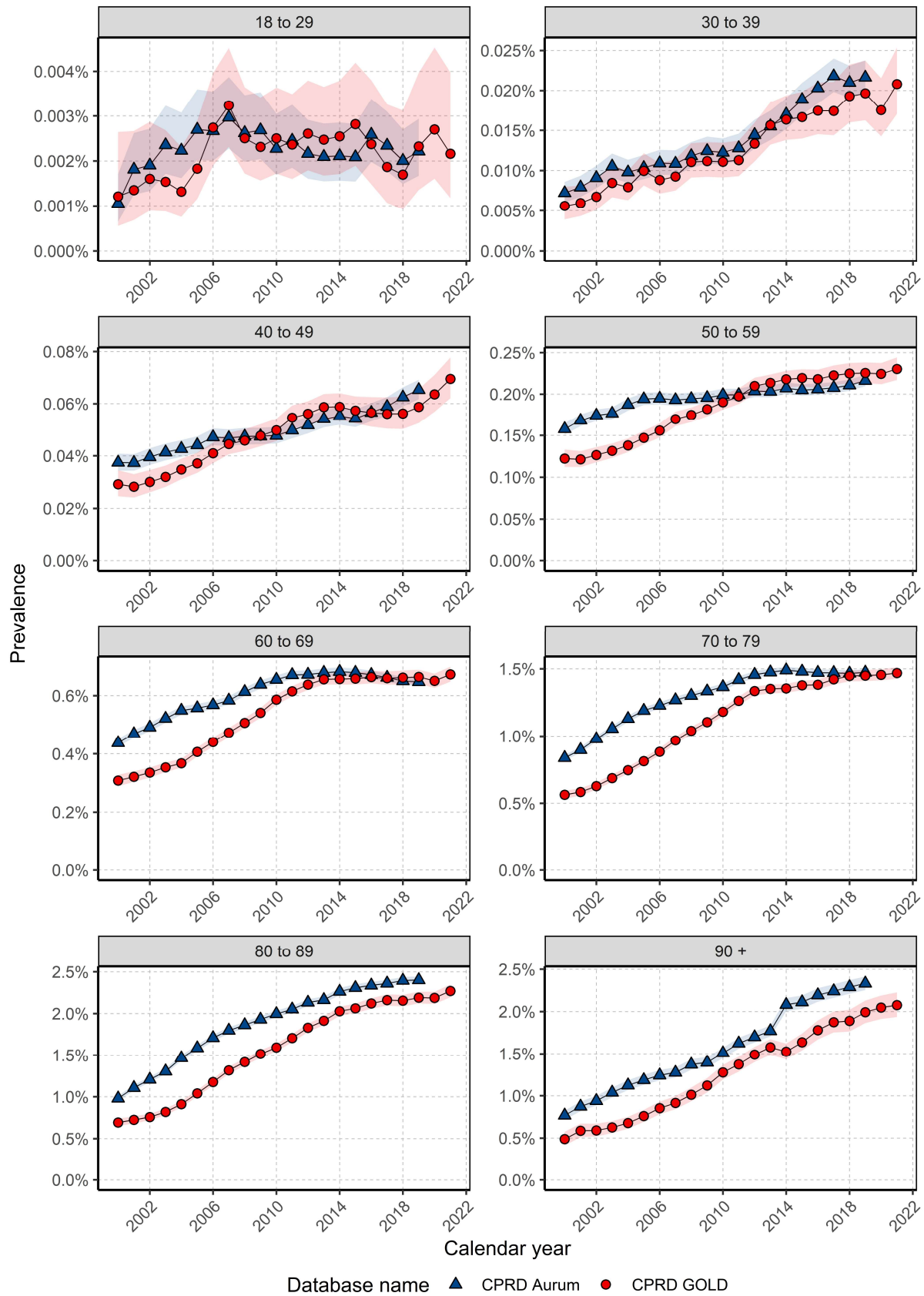


Figure 4: Annual prevalence from 2000 to 2021 stratified by database and age group.

Overall survival with stratification by sex, age and calendar year

In GOLD, there were 53,098 patients with 26,084 deaths (49.1%) over the study period whereas for Aurum, there were 98,569 patients with 48,894 deaths (49.6%). Both databases had similar median follow-up time of 2.4-2.5 years and median survival of 5.44-5.54 years.

For GOLD, survival after one, five and ten years from diagnosis was 78.3%, 51.4% and 38.5% for the whole population. Females had higher median survival compared to males (Supplement S7) with median survival times for females of 6.0 (5.8 to 6.3) years and 5.1 (4.9 to 5.3) years for males with similar results obtained in Aurum (males 5.2 years; females 6.13 years). Regarding short- and long-term survival, both sexes had similar one year survival of around 79%, whereas males had lower survival at five and ten years compared to females with survival of 50.3% and 52.8% at five years and 36.1% and 40.7% at ten years for males and females respectively. Similar results were also obtained in Aurum (Supplement S8).

When stratifying by age group, median survival decreased with age for both databases. Median survival decreased in GOLD from 15.6 years in those aged 50-59 years to 0.96 years for those over 90 years (Supplement S9). Survival at one, five, and ten years, generally decreased with increasing age from 50 years and older with the lowest survival observed in those aged 90 years and older (Supplement S10). Short term survival (one year) was between 86-90% for those aged 18-49 years whereas was 49-51% in those aged 90 years and older. Long term survival at five years decreased from 50 years of age and survival at ten-year decreased with increasing age from 30 years of age across both databases.

To investigate if survival rates have changed over time, we stratified by calendar time of cancer diagnosis in five-year age windows for GOLD. Figure 5 shows the KM survival curve stratified by sex and calendar year. In GOLD one- and five-year survival slightly improved in those diagnosed in 2000 to 2004 to those diagnosed in 2015 to 2019, increasing from 77.1% (76.2-78.0) to 78.9% (78.0-79.67) for one year survival and from 48.9% (47.8-50.0) to 52.0% (50.9-53.2) for five-year survival, respectively. Stratification by sex showed improvements in five-year survival over calendar time for both sexes, but not for one year survival (Supplement S11). Stratification by different age groups showed only in those aged 60-69 years showed improvements in both one-year (80.7% [79.0-82.5] to 85.8% [84.4-97.2]) and five-year survival (55.5% [53.3-57.8] to 64.1% [61.9-66.4]) comparing those diagnosed in 2000-2004 with those diagnosed in 2015-2019 with similar results for both males and females. There was no difference in survival for those diagnosed in 2020-2021 to previous calendar year groups.

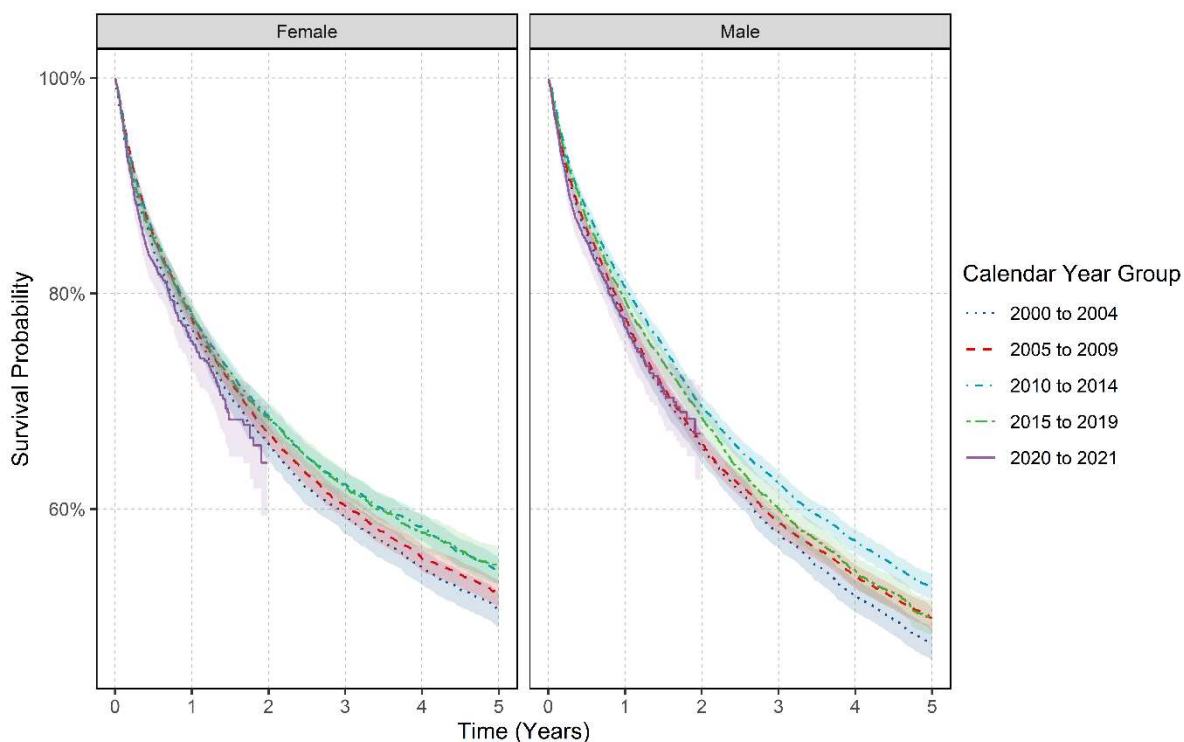


Figure 5: Kaplan-Meier survival curve of colorectal cancer stratified by sex and calendar year of diagnosis.

DISCUSSION

This paper provides a comprehensive study of the trends of CRC incidence, prevalence, and survival in the UK. Incidence increased until 2011 and prevalence steadily increased over the study period, with a relevant increase in those younger than 50 years of age. Survival slightly improved from 2000 to 2019 particularly in those aged 60 to 69 years.

Our estimates of incidence rates of CRC are in line with national statistics². Similar to previous literature, we also observed higher rates in men and older people^{3,21,22}. Prevalence also increased over the study period, with age group differences, likely due to the impact of screening programmes, as we observed a stabilisation during the last decade of the study in the targeted age groups²³.

We observed four main trends on incidence during the study period. First, a progressive increase after 2006, reaching its peak in 2011 and decreasing the subsequent years, with a sharp rise in 2021, likely driven by deferred diagnosis after the pandemic²⁴. Earlier trends could be associated with the introduction of the National Bowel Screening Programme in 2006²⁵. The detection of early-stage asymptomatic cases could explain an increase in IRs with a subsequent reduction due to endoscopic resection of adenomas^{7,26}. This pattern had already been described previously^{12,13}. A population-based study in the Netherlands showed an initial increase in incidence after the start of the screening programme, especially in early-stage CRC incidence, followed by a decrease in both early- and advanced-stage CRC incidence²⁷.

Third, we observed a worrying increase in cases in those under 50 years of age, which has also been observed in many other countries^{4,12}. The reasons for this increase in early-onset colorectal cancer (EOCRC) are unclear. EOCRC appears to have differential clinical and

pathological characteristics²⁸. One possible explanation is birth cohort effects^{29,30}, where those born after 1960's were exposed to modifiable risk factors such as unhealthy Western diets, sedentary behaviour, and lower physical activity^{29,31}. Interestingly, we did not observe sex differences in this age group, suggesting these exposures are not sex specific^{3,30,32}. On the other hand, the rising incidence of CRC in younger age groups has led to some countries, including the UK, to lower the bowel cancer screening age eligibility⁹. Notably, a decline in young-onset CRC is observed in those countries with a lowered age for CRC screening eligibility (Austria and Italy), while an increasing incidence is observed in other nine high-income countries (Australia, Canada, Denmark, Germany, New Zealand, Slovenia, Sweden, UK and USA)³⁰.

Fourth, the IRs also increased over the study period in nonagenarians. This could be due to improvements in life expectancy³³. Also, this age group does not benefit of routine cancer screening, as it is not recommended over 85 years of age³³. The diagnostic and treatment in this age group is challenging, as colonoscopy is an invasive procedure associated with increased risks and incomplete examinations in the elderly or frail patients³⁴. On the other hand, several studies point out that the age itself should not be a relevant criterion and suggest geriatric assessment to optimise cancer care pathways as well as including older patients in clinical trials^{35,36}. Notwithstanding the above, a north English population-based study found that age remains a major factor in treatment decisions in the UK³⁷.

Sex differences for CRC incidence and survival have previously been described^{3,6,22,38}. We observed higher IRs and worse overall survival in males, except for EO CRC. Negligible differences in age-standardized survival among men and women had been previously described in England²², while other studies had reported better overall and cancer-specific survival in women than men as well as better recurrence-free survival³⁹. These sex differences can be explained by several modifiable and non-modifiable factors. Males have higher exposure to environmental risk factors (obesity, unhealthy diet, and alcohol consumption), and higher vulnerability to them^{3,6,40}. Men also have lower uptake of CRC screening, although in England a higher proportion of CRCs in males are diagnosed via the screening programme²². This could be explained by the fact that women are more prone to have proximal and more aggressive cancers and because gFOBT/FIT screening is less effective in women²². On the other hand, sexual dimorphisms also could contribute to these differences⁶.

Short- and long-term survival results are in line with national statistics⁴¹ and with previous literature¹⁴. Median survival decreased with age as expected, which can be explained by frailty and comorbidities⁴².

Despite the improvement in therapeutic management in recent years, we only observed a modest impact on survival. Moreover, when stratifying by age, only those aged 60-69 years had consistent improvements, which could reflect the effects of the screening in this age group¹⁰. Survival rates in the UK have already been assessed in other studies. Compared to other European countries with similar health systems and populations, overall survival rates are worse in the UK, particularly for elderly people^{14,15}. Careful consideration is necessary when interpreting these comparisons, since these studies are based on cancer registries and there are intercountry differences in data collection, potentially introducing bias⁴³. Greater efforts are being made to understand the extensive heterogeneity of CRC to avoid the current "one-fits-for-all" therapeutic strategy, with emerging research moving closer to personalized treatments based on the molecular characteristics of the tumour and enhancing risk stratification, chemotherapy monitoring, and early relapse detection^{44,45}. Once this research

can be applied in clinical practice, it is expected that survival rates will enhance in the coming years.

This study has many strengths. First, we used two large primary care databases covering a large proportion of the UK, including England and the devolved nations ^{46,47}. The similarity between the results from both databases proves the robustness of our findings. The sharp increase in incidence at the start of the study period between 2000-2004 could be explained by the introduction of the Quality and Outcomes Framework (QOF) in 2004 which encouraged general practitioners to record all new cases ⁴⁸. Second, our study used a complete representative population-based database for the assessment of both numerators and denominators in the estimated incidence and prevalence. In contrast, cancer registry studies extrapolate the registry data to the whole population using national statistics which could potentially introduce bias ^{49,50}. Third, the high validity of mortality data with over 98% accuracy⁵¹ allowed us to examine the impact of calendar time on overall survival.

Our study also had limitations. Firstly, we used primary care data without linkage to cancer registry, potentially leading to misclassification and delayed recording of diagnoses. However, previous validation studies have shown high accuracy and completeness of cancer diagnoses in UK primary care records ⁵². Secondly, our use of primary care records precluded us from studying tumour histology, genetic mutations, staging or cancer therapies which can all impact survival. Therefore, our survival estimates may overestimate survival in those with higher staging ¹⁰ as well those with specific genetic mutations such as KRAS and BRAF ^{53,54}. Other factors such as socio-economic status and ethnicity could also result in different values for incidence, prevalence, and survival ⁵⁵.

CONCLUSION

In conclusion, our study shows a reduction in CRC incidence over the years mostly driven by those aged 60-79 years, in whom consistent small improvements in survival are also observed, which could be associated with screening. All this leads to an overall increase in CRC prevalence, which will require increased healthcare resources. On the other hand, a worryingly increase in the incidence of EOCRC raises the question whether lowering the screening age could contribute to better outcomes in these patients. It also reflects the necessity of a better understanding of the biology of young-onset CRC and associated risk factors, which will facilitate and expand risk-stratified screening in the future. Finally, despite advancements in therapeutic management, we only found modest improvements in survival which can be explained by the convergence of multiple factors and underscores the need for further research into the management of CRC.

ADDITIONAL INFORMATION

ACKNOWLEDGEMENTS

None.

AUTHOR CONTRIBUTIONS

All authors were involved in the study conception and design, interpretation of the results, and the preparation of the manuscript. AD and WYM implemented the data curation, data harmonisation, data quality tests and assessment. DN carried out data analysis for the

manuscript. AWR developed the clinical code lists for CRC for this study. AG and AWR reviewed the clinical codelists used in this study. PPP, CGA and DN wrote the initial draft of the manuscript with DPA. DN, EB, AD, WYM and DPA had access to the CPRD data. All authors critically reviewed the final manuscript and gave consent for publication.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The use of CPRD data for this study was approved by the Independent Scientific Advisory Committee (22_001843). Informed consent was waived by competent authorities due to the anonymized nature of patient data.

CONSENT FOR PUBLICATION

Not applicable.

DATA AVAILABILITY

This study is based in part on data from the Clinical Practice Research Datalink (CPRD) obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. The data is provided by patients and collected by the NHS as part of their care and support. The interpretation and conclusions contained in this study are those of the author/s alone. Patient level data used in this study was obtained through an approved application to the CPRD (application number 22_001843) and is only available following an approval process to safeguard the confidentiality of patient data. Details on how to apply for data access can be found at <https://cprd.com/data-access>.

COMPETING INTERESTS

Professor Daniel Prieto-Alhambra research group has received research grants from the European Medicines Agency, from the Innovative Medicines Initiative, from Amgen, Chiesi, and from UCB Biopharma; and consultancy or speaker fees (paid to his department) from Astellas, Amgen, Astra Zeneca, and UCB Biopharma. All other authors declare no conflicts of interest.

FUNDING

This activity under the European Health Data & Evidence Network (EHDEN) has received funding from the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking under grant agreement No 806968. IMI2 receives support from the European Union's Horizon 2020 research and innovation programme and European Federation of Pharmaceutical Industries and Associations (EFPIA). The sponsors of the study did not have any involvement in the writing of the manuscript or the decision to submit it for publication. Additionally, there was partial support from the Oxford NIHR Biomedical Research Centre. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021 May;71(3):209–49.
2. Cancer Research UK [Internet]. Available from: <https://www.cancerresearchuk.org>
3. Keum NN, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol*. 2019 Dec 1;16(12):713–32.
4. Exarchakou A, Donaldson LJ, Girardi F, Coleman MP. Colorectal cancer incidence among young adults in England: Trends by anatomical subsite and deprivation. *PLoS One*. 2019 Dec 1;14(12).
5. Abancens M, Bustos V, Harvey H, McBryan J, Harvey BJ. Sexual Dimorphism in Colon Cancer. *Front Oncol*. 2020;10(December):1–27.
6. Baraibar I, Ros J, Saoudi N, Salvà F, García A, Castells MR, et al. Sex and gender perspectives in colorectal cancer. *ESMO Open* [Internet]. 2023;8(2):101204. Available from: <https://doi.org/10.1016/j.esmoop.2023.101204>
7. Brenner H, Chen C. The colorectal cancer epidemic: challenges and opportunities for primary, secondary and tertiary prevention. Vol. 119, *British Journal of Cancer*. Nature Publishing Group; 2018. p. 785–92.
8. Navarro M, Nicolas A, Ferrandez A, Lanás A. Colorectal cancer population screening programs worldwide in 2016: An update. *World J Gastroenterol*. 2017;23(20):3632–42.
9. Public Health England. Bowel cancer screening: programme overview [Internet]. Available from: <https://www.gov.uk/guidance/bowel-cancer-screening-programme-overview>
10. Cardoso R, Guo F, Heisser T, De Schutter H, Van Damme N, Nilbert MC, et al. Overall and stage-specific survival of patients with screen-detected colorectal cancer in European countries: A population-based study in 9 countries. *The Lancet Regional Health - Europe*. 2022;21:1–11.
11. Soriano LC, Soriano-Gabarró M, García Rodríguez LA. Trends in the contemporary incidence of colorectal cancer and patient characteristics in the United Kingdom: A population-based cohort study using The Health Improvement Network. *BMC Cancer*. 2018 Apr 10;18(1).
12. Araghi M, Soerjomataram I, Bardot A, Ferlay J, Cabasag CJ, Morrison DS, et al. Changes in colorectal cancer incidence in seven high-income countries: a population-based study. *Lancet Gastroenterol Hepatol*. 2019 Jul 1;4(7):511–8.
13. Cardoso R, Guo F, Heisser T, Hackl M, Ihle P, De Schutter H, et al. Colorectal cancer incidence, mortality, and stage distribution in European countries in the colorectal cancer screening era: an international population-based study. *Lancet Oncol*. 2021 Jul 1;22(7):1002–13.
14. Holleczeck B, Rossi S, Domenic A, Innos K, Minicozzi P, Francisci S, et al. On-going improvement and persistent differences in the survival for patients with colon and rectum cancer across Europe 1999-2007 - Results from the EURO CARE-5 study. *Eur J Cancer*. 2015 Oct 1;51(15):2158–68.
15. Araghi M, Arnold M, Rutherford MJ, Guren MG, Cabasag CJ, Bardot A, et al. Colon and rectal cancer survival in seven high-income countries 2010-2014: Variation by age and stage at diagnosis (the ICBP SURVMARK-2 project). *Gut*. 2021 Jan 1;70(1):114–26.
16. Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *Journal of the American Medical Informatics Association*. 2015;22(3):553–64.

17. Schuemie M, Ryan P, Weaver J. Cohort diagnostics: diagnostics for OHDSI studies. 2020.
18. Eurostat Taskforce. Revision of the European Standard Population - Report of Eurostat's task force. 2013. [Internet]. [cited 2023 Aug 2]. Available from: <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-ra-13-028>
19. Raventós B, Català M, Du M, Guo Y, Black A, Inberg G, et al. IncidencePrevalence: An R package to calculate population-level incidence rates and prevalence using the OMOP common data model. *Pharmacoepidemiol Drug Saf.* 2024;33(1):1–11.
20. Therneau T. A Package for Survival Analysis in R. R package version 3.5-8 [Internet]. 2024. Available from: <https://cran.r-project.org/package=survival>
21. Howlader N, Noone A, Krapcho M, Miller D, Brest A, Yu M, et al. SEER Cancer Statistics Review, 1975-2017, National Cancer Institute. [Internet]. April 2020. Available from: https://seer.cancer.gov/csr/1975_2017/
22. White A, Ironmonger L, Steele RJC, Ormiston-smith N, Crawford C, Seims A. A review of sex-related differences in colorectal cancer incidence , screening uptake , routes to diagnosis , cancer stage and survival in the UK. 2018;1–11.
23. Id JB. Prevalence threshold (ϕe) and the geometry of screening curves. 2020;1–12.
24. Barclay N, Burkard T, Burn E, Delmestri A, Miquel Dominguez A, Golozar A, et al. The Impact of the COVID-19 Pandemic on Incidence and Short-Term Survival for Common Solid Tumours in the United Kingdom: A Cohort Analysis. *Clin Epidemiol.* 2024;Volume 16(June):417–29.
25. Bretthauer M, Løberg M, Wieszczy P, Kalager M, Emilsson L, Garborg K, et al. Effect of Colonoscopy Screening on Risks of Colorectal Cancer and Related Death. *New England Journal of Medicine.* 2022;387(17):1547–56.
26. Winawer SJ, Zauber AG, Ho M, O'Brien MJ, Gottlieb L, Sternberg SS, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *New England Journal of Medicine* [Internet]. 1993;329(27):1977–81. Available from: <http://content.nejm.org/cgi/content/abstract/329/14/977%5Cnhttp://www.nejm.org/doi/abs/10.1056/NEJM199309303291401>
27. Breekveldt ECH, Lansdorp-Vogelaar I, Toes-Zoutendijk E, Spaander MCW, Vuuren AJ van, Kemenade FJ van, et al. Colorectal cancer incidence, mortality, tumour characteristics, and treatment before and after introduction of the faecal immunochemical testing-based screening programme in the Netherlands: a population-based study. *Lancet Gastroenterol Hepatol.* 2022;7(1):60–8.
28. Burnett-Hartman AN, Lee JK, Demb J, Gupta S. An Update on the Epidemiology, Molecular Characterization, Diagnosis, and Screening Strategies for Early-Onset Colorectal Cancer. *Gastroenterology.* 2021 Mar 1;160(4):1041–9.
29. Murphy CC, Zaki TA. Changing epidemiology of colorectal cancer — birth cohort effects and emerging risk factors. 2024;21(January).
30. Siegel RL, Torre LA, Soerjomataram I, Hayes RB, Bray F, Weber TK, et al. Global patterns and trends in colorectal cancer incidence in young adults. *Gut.* 2019;68(12):2179–85.
31. Kim NH, Jung YS, Yang H joon, Park S kyung, Park JH, Park D II, et al. Prevalence of and Risk Factors for Colorectal Neoplasia in Asymptomatic Young Adults (20-39 Years Old). *Clinical Gastroenterology and Hepatology* [Internet]. 2019;17(1):115–22. Available from: <https://doi.org/10.1016/j.cgh.2018.07.011>
32. Patel P, De P. Trends in colorectal cancer incidence and related lifestyle risk factors in 15-49-years-olds in Canada, 1969-2010. *Cancer Epidemiol.* 2016;42:90–100.

33. DeSantis CE, Miller KD, Dale W, Mohile SG, Cohen HJ, Leach CR, et al. Cancer Statistics for Adults Aged 85 Years and Older, 2019. *CA Cancer J Clin*. 2019;(69):452–67.
34. Machlab S, Francia E, Mascort J, Garcia-Iglesias P, MEndive JM, Riba F, et al. Riesgos, indicaciones y aspectos técnicos de la colonoscopia en pacientes de edad avanzada o con fragilidad. Documento de posicionamiento de la Societat Catalana de Digestologia, la Societat Catalana de Geriatria i Gerontologia i la Societat Catalana de M. Gastroenterol Hepatol. 2024;47:107–17.
35. Nakayama Y, Kawamura H, Honda M, Takano Y, Takiguchi K, Kamiga T, et al. Benefit of intensive chemotherapy for elderly patients aged 80 years or older with metastatic colorectal cancer: a state-wide multicenter cohort study. *Int J Clin Oncol*. 2021 Jul 1;26(7):1248–56.
36. Ose I, Rosen AW, Bräuner K, Colov EBP, Christensen MG, Mashkooor M, et al. The association of postoperative morbidity and age on 5-year survival after colorectal surgery in the elderly population: a nationwide cohort study. *Colorectal Disease*. 2024;26(5):899–915.
37. Hayes L, Forrest L, Adams J, Hidajat M, Ben-shlomo Y, White M, et al. Age-related inequalities in colon cancer treatment persist over time: a population-based analysis. 2018;(December 2011):1–8.
38. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin*. 2020 May;70(3):145–64.
39. Samawi HH, Yin Y, Speers CH, Cheung WY. Sex Disparities in Outcomes of Early Stage Colorectal Cancer: A Population-Based Study. *Clin Colorectal Cancer*. 2024;17(4):e711–7.
40. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut*. 2023;(72):338–44.
41. Office For National Statistics [Internet]. Available from: <https://www.ons.gov.uk>
42. Wang M jin, Ping J, Li Y, Adell G, Arbman G, Nodin B. The prognostic factors and multiple biomarkers in young patients with colorectal cancer. *Sci Rep*. 2015;5(10645):1–11.
43. Souhami R. Are UK cancer cure rates worse than in most other European countries? *British Journal of General Practice*. 2010;81–2.
44. Nicolantonio F Di. Precision oncology in metastatic colorectal cancer — from biology to medicine. *Nat Rev Clin Oncol* [Internet]. 2021;18. Available from: <http://dx.doi.org/10.1038/s41571-021-00495-z>
45. Roazzi L, Patelli G, Bencardino KB, Amatu A, Bonazzina E, Tosi F, et al. Ongoing Clinical Trials and Future Research Scenarios of Circulating Tumor DNA for the Treatment of Metastatic Colorectal Cancer. 2024;
46. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Staa T van, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827–36.
47. Wolf A, Dedman D, Campbell J, Booth H, Lunn D, Chapman J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol*. 2019;48(6):1740-1740G.
48. Prescribing and Primary Care team Health and Social Care Information Centre, 'Quality and Outcomes Framework, Achievement, prevalence and exceptions data, 2011/12' [Internet]. Available from: <https://digital.nhs.uk/>
49. Swerdlow A. J. Cancer Registration in England and Wales: Some Aspects Relevant to Interpretation of the Data. *Journal of the Royal Statistical Society*. 1986;149(2):146–60.

50. Alsadhan N, Almaiman A, Pujades-Rodriguez M, Brennan C, Shuweihdi F, Alhurishi SA, et al. A systematic review of methods to estimate colorectal cancer incidence using population-based cancer registries. *BMC Med Res Methodol* [Internet]. 2022;22(1):1–15. Available from: <https://doi.org/10.1186/s12874-022-01632-7>
51. Gallagher AM, Dedman D, Padmanabhan S, Leufkens HGM, de Vries F. The accuracy of date of death recording in the Clinical Practice Research Datalink GOLD database in England compared with the Office for National Statistics death registrations. *Pharmacoepidemiol Drug Saf*. 2019;28(5):563–9.
52. Strongman H, Williams R, Bhaskaran K. What are the implications of using individual and combined sources of routinely collected data to identify and characterise incident site-specific cancers? a concordance and validation study using linked English electronic health records data. *BMJ Open*. 2020;10(8):1–12.
53. Meng M, Zhong K, Jiang T, Liu Z, Kwan HY, Su T. The current understanding on the impact of KRAS on colorectal cancer. *Biomedicine and Pharmacotherapy* [Internet]. 2021;140(February):111717. Available from: <https://doi.org/10.1016/j.biopha.2021.111717>
54. Xu T, Li J, Wang Z, Zhang X, Zhou J, Lu Z, et al. Real-world treatment and outcomes of patients with metastatic BRAF mutant colorectal cancer. *Cancer Med*. 2023;12(9):10473–84.
55. Aarts MJ, Lemmens VEPP, Louwman MWJ, Kunst AE, Coebergh JWW. Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. *Eur J Cancer* [Internet]. 2010;46(15):2681–95. Available from: <http://dx.doi.org/10.1016/j.ejca.2010.04.026>