Cohort profile: The National Colorectal Cancer Cohort (NCRCC) study in China

Yingshuang Zhu, Yeting Hu, Xiangxing Kong, Qian Xiao, Zhizhong Pan, Zhaoxu Zheng, Ye Wei, Wang Ziqiang, Da Wang, Jiaqi Chen, Kun Chen, Shu Zheng, Meilin Wang, Xifeng Wu, Kefeng Ding, NCRCC Research Group

ABSTRACT

Purpose The National Colorectal Cancer Cohort (NCRCC) study aims to specifically assess risk factors and biomarkers related to endpoints across the colorectal cancer continuum from the aetiology through survivorship.

Participants The NCRCC study includes the Colorectal Cancer Screening Cohort (CRCSC), which recruited individuals who were at high risk of CRC between 2016 and 2020 and Colorectal Cancer Patients Cohort (CRCPC), which recruited newly diagnosed patients with CRC between 2015 and 2020. Data collection was based on questionnaires and abstraction from electronic medical record. Items included demographic and lifestyle factors, clinical information, survivorship endpoints and other information. Multiple biospecimens including blood, tissue and urine samples were collected. Participants in CRCSC were followed by a combination of periodic survey every 5 years and annual linkage with regional or national cancer and death registries for at least 10 years. In CRCPC, follow-up was conducted with both active and passive approaches at 6, 12, 18, 24, 36, 48 and 60 months after surgery.

Findings to date A total of 19,377 participants and 15,551 patients with CRC were recruited in CRCSC and in CRCPC, respectively. In CRCSC, 48.0% were men, and the average age of participants at enrolment was 58.7±8.3 years. In CRCPC, 61.4% were men, and the average age was 60.3±12.3 years with 18.9% of participants under 50 years of age.

Future plans Longitudinal data and biospecimens will continue to be collected. Based on the cohorts, several studies to assess risk factors and biomarkers for CRC or its survivorship will be conducted, ultimately providing research evidence from Chinese population and optimising evidence-based guidelines across the CRC continuum.

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related deaths worldwide. In China, the CRC incidence and mortality have rapidly increased in recent years. In parallel, the number of CRC survivors continues to increase, which is partly due to wider participation in CRC screening and personalised comprehensive treatment. Therefore, recognising that prognosis in cancer survivors has tremendous associations with early diagnosis, treatment and health behaviours, evidence-based guidelines are urgently required to improve these processes.

Studies of lifestyle factors, genetic and molecular profiles can contribute to discovery of novel and powerful predictors of early screening, clinical outcomes and survival. However, such studies require large sample size with comprehensive exposure data, well-annotated clinical outcomes, multiple types of biospecimens and high-quality follow-up information. Compared with the well-known established CRC cohorts, there are few large prospective cohorts that combine a comprehensive set of behavioural factors, together with detailed assessments of clinical outcomes and survivorship endpoints in survivors. Additionally, all these established CRC patient cohorts provided sample size with less than 10,000 and the study populations were restricted in developed countries. However, patterns of exposure vary markedly between developed countries and developing countries, especially in diet and lifestyle. Therefore, there is a critical need for a large CRC cohort capable of identifying potentially modifiable cancer.
factors and novel biomarkers for screening and clinical outcomes, and assessing survivorship related to CRC treatments, especially in developing countries.

With an unprecedented economic development in recent years, the whole landscape of lifestyle behaviours in Chinese population has been largely changed. Recent researches have shown that the increasing CRC incidence and mortality were mainly due to unhealthy lifestyle changes, such as obesity, Western dietary habits, insufficient exercise, smoking and excessive alcohol drinking. However, to date, still little is known about the effects of changed lifestyle on incidence and prognosis of CRC among the Chinese population. Meanwhile, existing diagnosis and treatment guidelines of CRC have limited research evidence from the Chinese population, and the clinical outcomes of comprehensive treatments to CRC in developing countries require further verification by large cohort studies. Therefore, it is desirable to develop a cohort with large sample size in China to support a series of well-powered studies and provide evidence to improve guidelines of CRC screening, diagnosis, treatment and management for the Chinese population.

RESEARCH QUESTIONS
The National Colorectal Cancer Cohort (NCRCC) study aims to recruit patients with CRC and individuals who have one or more high-risk factors of CRC. It was developed specifically to support valuable studies involving aetiological analysis, early screening and diagnosis, clinical outcomes and survivorship in well-characterised Chinese population. Specific research purposes and contents of the NCRCC study were:

1) To integratively investigate lifestyle and genetic risk factors for CRC to develop risk prediction models.

We will assess the specific exposure spectrum in Chinese population, such as diet, alcohol drinking, smoking, physical activity, body fatness and aspirin use and investigate the association between these lifestyle factors and the risk of CRC. Single nucleotide polymorphism genotyping will be conducted to estimate the association between genetic variants and CRC risk. Since CRC is a complex disease that results from both genetic and lifestyle factors, interactions between lifestyle factors and genetic variants will also be evaluated. Finally, lifestyle and genetic risk factors will be combined to develop risk prediction models.

2) To discover and validate new biomarkers for early screening and diagnosis, and prognosis.

We will identify new biomarkers (e.g., metagenomic, metabolomic, genetic, epigenetic and proteomic biomarkers) for early screening and diagnosis. Previous studies have shown that gut microbiota contribute to the development of CRC and have potential to be biomarkers for CRC and adenoma diagnosis. At the present stage, we mainly focus on gut microbiota and their metabolites. For example, circulating bacterial DNA has been reported to discriminate other cancers (e.g., prostate cancer and lung cancer) from healthy controls; however, it remains unexplored in patients with CRC. We will characterise alterations of bacterial DNA and their metabolites in plasma samples and identify bacteria-related biomarkers for CRC early detection as well as prognosis.

3) To study the treatment response and survival disparities among patients and identify contributors to the observed disparities.

We will assess the treatment response of different therapies, such as neoadjuvant therapy, immunotherapy and identify the impact factors or applicable indicators of these therapies. For example, the gut microbiota and immune-related markers associated with the response to neoadjuvant chemoradiotherapy will be studied. In addition, the association of surgical treatments (e.g., complete mesocolic excision) and related operations with survival will also be investigated.

4) To provide evidence-based recommendations for lifestyle behaviour changes to reduce the risk of CRC recurrence and therapy-induced adverse effect and improve the quality of life in survivors.

We will assess the lifestyle behaviour including diet, physical activity and medication use (e.g., aspirin, statins and metformin) by questionnaires at diagnosis and during the following years and estimate the association between these modifiable factors and the risk of CRC recurrence among cancer survivors. Additionally, many patients with CRC who received treatments (e.g., chemoradiotherapy) suffer from short and long-term adverse effects (e.g., second primary cancer). Therefore, in this section, we will also focus on the impact of lifestyle behaviours on late effect of treatment and provide evidence for healthy lifestyle recommendation to improve the quality of life in survivors.

COHORT DESCRIPTION
The NCRCC study includes two subcohorts: the Colorectal Cancer Screening Cohort (CRCSC), which is a community-based cohort recruiting participants who are identified as individuals at high risk of CRC and subsequently perform colonoscopy examinations, and the Colorectal Cancer Patients Cohort (CRCPC), in which newly diagnosed patients with CRC are recruited.

Colorectal Cancer Screening Cohort
The CRCSC was first launched in Jiashan County, Zhejiang Province, by the Second Affiliated Hospital of Zhejiang University, in January 2016. All individuals aged 40–74 years old who were permanent residents of Jiashan County and without history of CRC were eligible for CRC screening. The details of the screening were described in our previous study. Briefly, individuals were interviewed with a high-risk factor questionnaire (HRFQ) and submitted two faecal samples for faecal immunochemical tests (FITs). Those with a history of colorectal polyps or cancers other than CRC, or family history of CRC, or two or more of the following items: (1) chronic diarrhoea, (2) chronic constipation, (3) stressful life events, (4) mucous...
Table 1 Characteristics of participants at enrolment in NCRCC study.

| Characteristics          | CRCSC N=19 377 (%) | CRCPC N=15 551 (%) |
|--------------------------|--------------------|--------------------|
| Gender                   |                    |                    |
| Men                      | 48.2               | 61.4               |
| Women                    | 51.8               | 38.6               |
| Age (year)               | 58.7±8.3           | 60.3±12.3          |
| <50                      | 15.1               | 18.9               |
| 50–65                    | 56.8               | 42.9               |
| ≥65                      | 28.1               | 38.2               |
| Education                |                    |                    |
| No formal education      | 30.4               | 21.8               |
| Elementary               | 33.2               | 26.3               |
| Middle school            | 30.7               | 25.0               |
| High/technical school    | 4.7                | 12.9               |
| College or above         | 1.0                | 14.0               |
| Family history of cancer | 29.9               | 26.6               |
| BMI (kg/m²)              | 23.4±3.1           | 23.2±3.3           |
| Normal weight (18.5–24)  | 55.39              | 53.8               |
| Underweight (<18.5)      | 4.4                | 6.9                |
| Overweight (24–27.9)     | 33.2               | 31.8               |
| Obesity (≥28)            | 7.1                | 7.5                |

CRCPC, Colorectal Cancer Patients Cohort; CRCSC, Colorectal Cancer Screening Cohort; NCRCC, National Colorectal Cancer Cohort.

for the study. Patients with a history of CRC or any mental condition that makes it impossible to provide informed consent or complete questionnaires were excluded. Each subject in the CRCPC was interviewed by a trained general physician or nurse in the local region. A baseline questionnaire was administered and the structured data of clinical information were abstracted from the electronic medical record (EMR). Standardised recruitment protocols were implemented at each research site. As of June 2020, 15 551 participants were enrolled in the CRCPC. The majority (61.4%) of participants were men, and 18.9% were under 50 years of age (table 1). A total of 73.1% received less than a high or technical school education.

The study had been registered at www.ClinicalTrials.gov. Detailed data and biospecimen collection procedures of the study are shown in figure 1.

How often have been followed up?

CRCSC
Participants in the CRCSC were followed up by a combination of periodic CRC screening and linkage with regional or national cancer and death registry systems, and the follow-up would last for at least 10 years. The new CRC cases were identified by annually checking the files of cancer registry system, which included information on cancer location, histology and diagnosis date. The information on vital status was updated annually via the death registry system. In addition, each subject enrolled in the cohort would update their exposure and health status every 5 years during periodic CRC screening.

CRCPC
Follow-up for the CRCPC was conducted with both active and passive approaches and would last for at least 5 years. Participants were requested to come back to the hospital to check the health status at 6, 12, 18, 24, 36, 48 and 60 months after surgery. For patients undergoing adjuvant chemotherapy or (and) radiotherapy, reviews of medical records and exposure assessment were also performed at the end of each cycle. The research centres collected information on updated health status and behaviours and related specimens, including blood and urine samples. Computer matches with the cohort follow-up files identified patients who have presented for a recent appointment versus those who have not visited during each follow-up period. For the latter group, a phone call was sent to interview about the patient’s health status and lifestyle behaviours. CRC recurrence/metastasis was ascertained by carefully reviewing the patient’s medical record, including pathology and imaging reports. For patients who had not regularly visited the recruiting hospital during the follow-up period and shown any signs of cancer recurrence or metastasis, a review of outside medical records would be performed. Finally, for those lost to follow-up with the above two approaches, information on vital status was retrieved by linkage with a vital registration system.
What has been measured?
An NCRCC information system was developed to jointly record baseline data, clinical outcomes, follow-up data and biospecimen information and provide a platform for sharing data resources. Baseline data from the electronic questionnaire can be interconnected and submitted to the information system through the internet. Clinical outcomes of all participants were abstracted in a standardised form and submitted to the system. Biospecimens were stored at each research site but with centralised management of biospecimen information by the Second Affiliated Hospital of Zhejiang University.

Data collection
A summary of the information collected in the NCRCC study is presented in table 2.

Baseline information
In the CRCSC, each individual was administered an electronic baseline questionnaire at recruitment. Items included demographics, lifestyle factors, diet, medication and dietary supplements use; for women, information on hormone replacement therapy, menopausal status and menstrual and reproductive history were also collected (table 2).

In the CRCPC, as part of the institutional registration process, all newly diagnosed patients were interviewed before treatment to complete baseline questionnaires, which include information on demographics (eg, gender, education, marital status and race), medical history, family history and medication use (eg, aspirin, statins, metformin). Since July 2018, an electronic questionnaire with more comprehensive baseline information (table 2) was first implemented at the Second Affiliated Hospital of Zhejiang University and would be gradually implemented at the remaining research sites. As of June 2020, 2122 records were collected using this tool.

For both cohorts, some specific exposures of Chinese populations, such as yellow rice wine, Chinese tea and diet, were investigated. Body weight, height, waist and hip circumference were measured by trained researchers at recruitment in both cohorts.

Clinical outcomes
In the CRCSC, for participants who had FIT or colonoscopy examination, the results of FIT or colonoscopy were collected. For participants with colorectal lesions after screening, information on pathological features and the following treatment were collected (table 2). In the
Table 2  Baseline, clinical and follow-up information collected in the NCRCC study

| Cohort         | Measurements                                                                                                                                 |
|----------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Baseline information | CRCSC, CRCPC Demographic (eg, gender, education, marital status and race), family history of cancer, medical history, lifestyle factors (eg, tobacco and alcohol use history, physical activity, shift work, and sleep), diet (vegetables, red meat, white meat, fruit, soy food, milk, et al), medication use (eg, aspirin, statins, metformin, antibiotics), dietary supplement use (multivitamins, calcium), and for women, hormone replacement therapy, menopausal status, menstrual and reproductive history. Body weight, height, waist and hip circumference. |
| Clinical outcome | CRCSC Diagnosis date, results of FIT and colonoscopy (number, location, size), pathological features and treatment.                                                                                  |
|                 | CRCPC Diagnosis date, AJCC stage, tumour location, pathological features (eg, histology, cell differentiation, lymphovascular or perineural invasion and number of positive lymph nodes), molecular features (MSI, KRAS, BRAF), imaging report (CT, colonoscopy), results of laboratory (CEA, CA199, et al), information on surgical procedures, surgical complications, treatment (ingredient, modality, timing and dose), tumour response and toxicities. |
| Follow-up data  | CRCSC Lifestyle factors (eg, tobacco and alcohol use history, physical activity, shift work, and sleep), diet, medication use (eg, aspirin, statins, metformin, antibiotics), dietary supplement use (multivitamins, calcium), patients’ health (diagnosis of colorectal cancer) and vital status (survivorship, date and cause of death). |
|                 | CRCPC Lifestyle factors (tobacco and alcohol use, physical activity), patients’ health (recurrence/metastasis, treatment hospital, diagnosis of second primary cancer) and vital status (survivorship, date and cause of death) |

AJCC, American Joint Committee on Cancer; CA199, carbohydrate antigen 19; CEA, carcinoembryonic antigen; CRCPC, Colorectal Cancer Patients Cohort; CRCSC, Colorectal Cancer Screening Cohort; FIT, faecal immunochemical test; MSI, microsatellite instability.

CRCPC, detailed information on clinical outcomes was standardised and abstracted from the EMR (table 2).

Follow-up data
In the CRCSC, the diagnosis of CRC and vital status were collected through cancer and death registry systems. Detailed information on exposure, including lifestyle factors, diet, medication and dietary supplements, would be updated at periodic CRC screening (table 2).

In the CRCPC, information on lifestyle factors (eg, tobacco and alcohol use, physical activity), patients’ health (recurrence/metastasis, treatment hospital, diagnosis of second primary cancer) and vital status was collected at each follow-up.

Furthermore, we plan to collect more epidemiological data (eg, diet, medication and dietary supplement use) during the follow-up period to explore the association between postdiagnosis lifestyle behaviours and survival. In addition, data on quality of life (European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-C30) were taken into consideration in 2020.

Biospecimen collection
Biospecimens, including blood and tissue samples, were collected, processed and stored according to the same standardised protocols at all research sites. In the CRCSC, each subject provided two 5 mL blood samples at recruitment. Biopsy tissue samples, which included polyp or tumour tissue and normal mucosa, were collected from participants who had lesions. In the CRCPC, newly recruited participants provided two 5 mL blood samples before treatment and immediately after surgery, respectively. All blood samples were kept at 4°C, processed into plasma, serum and buffy coat within 6 hours and stored at −80°C. Tissue samples that included tumour tissue and normal mucosa (both adjacent and distant to the tumour) were collected as both formalin-fixed paraffin-embedded (FFPE) and fresh frozen status. Fresh frozen tissue samples were immediately frozen in liquid nitrogen after excision and delivered to be stored at −80°C within 30 minutes to minimise the loss of RNA quality. For patients who had a relapse or metastasis and were treated in the recruiting hospital during the follow-up period, tissue samples were collected. In addition, 5 mL blood and urine samples at each clinic visit during follow-up period were collected in the Second Affiliated Hospital of Zhejiang University.

Biospecimen collection rates differed by sample type, time point and research centre. In the CRCSC, more than 80% of participants had blood samples. In the CRCPC, blood samples at baseline are currently available for 62.5% of participants. Fresh frozen tissues were collected from approximately 30% of patients with CRC undergoing surgery, while FFPE tissues were generally available.

Patient and public involvement
Patients and public were not involved in the development of this study.
FINDINGS TO DATE

Until June 2020, nearly 35,000 participants have been enrolled in the NCRCC study, with 19,577 participants in the CRCSC and 15,551 patients in the CRCPC. The demographic and selected baseline characteristics of the recruited NCRCC participants are shown in tables 1 and 3. Some selected clinical characteristics of CRCPC participants are given in table 4.

In the CRCSC, the prevalence of overweight/obesity (BMI ≥24 kg/m²) was 40.3%, and that of obesity (BMI ≥28 kg/m²) was 7.1%. A total of 3.6% patients had ever smoked and 28.2% currently smoked. Among the smokers, 14.7% smoked more than 20 cigarettes per day, and the average pack-years of smoking was 34.2±18.1 years. A total of 1.5% and 20.4% of participants reported having ever drunk and currently drinking alcohol, respectively. Among the drinkers, 14.0% were very high-risk drinkers (more than 100 g of alcohol consumption per day for men and 60 g for women).

A total of 33.8% of participants had ever drunk/currently drank Chinese tea and 13.4% engaged in regular exercise. At the baseline survey, 34.8%, 7.6% and 5.8% of the participants reported a history of hypertension, diabetes and vascular disease, respectively.

In the CRCPC, 16.6% had ever smoked, and 24.1% currently smoked. Among the smokers, 9.0% smoked more than 20 cigarettes per day, and the average pack-years of smoking was 31.1±19.2 years. A total of 10.8% had ever drunk and currently drank alcohol, respectively. Among the drinkers, 19.0% had very high-risk drinking (more than 100 g of alcohol consumption per day for men and 60 g for women). A total of 33.8% of participants had ever drunk/currently drank Chinese tea and 13.4% engaged in regular exercise. At the baseline survey, 34.8%, 7.6% and 5.8% of the participants reported a history of hypertension, diabetes and vascular disease, respectively.

In the CRCPC, 16.6% had ever smoked, and 24.1% currently smoked. Among the smokers, 9.0% smoked more than 20 cigarettes per day, and the average pack-years of smoking was 31.1±19.2 years. A total of 10.8% had ever drunk and currently drank alcohol, respectively. Among the drinkers, 19.0% had very high-risk drinking (more than 100 g of alcohol consumption per day for men and 60 g for women). A total of 33.8% of participants had ever drunk/currently drank Chinese tea and 13.4% engaged in regular exercise. At the baseline survey, 34.8%, 7.6% and 5.8% of the participants reported a history of hypertension, diabetes and vascular disease, respectively.

In the CRCPC, 16.6% had ever smoked, and 24.1% currently smoked. Among the smokers, 9.0% smoked more than 20 cigarettes per day, and the average pack-years of smoking was 31.1±19.2 years. A total of 10.8% had ever drunk and currently drank alcohol, respectively. Among the drinkers, 19.0% had very high-risk drinking (more than 100 g of alcohol consumption per day for men and 60 g for women). A total of 33.8% of participants had ever drunk/currently drank Chinese tea and 13.4% engaged in regular exercise. At the baseline survey, 34.8%, 7.6% and 5.8% of the participants reported a history of hypertension, diabetes and vascular disease, respectively.

**Table 3** Selected baseline characteristics of participants at enrolment in NCRCC study

| Selected characteristics | CRCSC (%) | CRCPC (%) |
|--------------------------|-----------|-----------|
| Income (Yuan per year)   | N=19377   | N=2122*   |
| Less than 32000          | 60.8      | 38.8      |
| 32 000–50 000            | 26.4      | 33.7      |
| More than 50 000         | 12.7      | 27.5      |
| WHR (mean±SD)            | 0.91±0.05 | 0.92±0.06 |
| Waist circumference (cm) | 85.1±8.1  | 81.0±8.8  |
| Tobacco                  |           |           |
| Ever smoked              | 3.6       | 16.6      |
| Currently smoking        | 28.2      | 24.1      |
| 1–14 cigarettes per day  | 27.5      | 35.3      |
| 15–20 cigarettes per day | 57.8      | 55.6      |
| >20 cigarettes per day   | 14.7      | 9.0       |
| Pack-years of smoking (mean±SD)| 34.2±18.1 | 31.1±19.2 |
| Alcohol                  |           |           |
| Ever drank               | 1.5       | 10.8      |
| Currently drinking       | 20.4      | 23.8      |
| Low risk drinking†       | 30.3      | 44.9      |
| Medium risk drinking‡    | 28.1      | 21.1      |
| High risk drinking§      | 27.7      | 15.0      |
| Very high risk drinking¶ | 14.0      | 19.0      |
| Green tea                |           |           |
| Former drinking          | 0.2       | 0.5       |
| Currently drinking       | 33.6      | 26.6      |
| Red meat intake (≥4 times per week) | 51.9 | 39.2 |
| Vegetable intake (≥6 times per week) | 94.5 | 19.5 |
| Fruit intake (≥4 times per week) | 25.0 | 28.0 |
| Regular exercise         | 13.4      | 34.1      |
| Sleep time (hour, mean±SD)| 8.4±1.2  | 8.7±1.0   |
| Aspirin use              | 4.9       | 4.5       |
| Statin use               | –         | 4.7       |
| Metformin use            | –         | 4.1       |
| Hypertension             | 34.8      | 35.8      |
| Diabetes                 | 7.6       | 12.1      |
| Vascular disease         | 5.8       | 7.4       |

*2122 participants who received an electronic version questionnaire.
140 g of alcohol consumption per day or less for men and 20 g for women.
41–60 g of alcohol consumption per day for men and 21–40 g for women.
≥61–100 g of alcohol consumption per day for men and 41–60 g for women.
†More than 100 g of alcohol consumption per day for men and 60 g for women.
CRCPC, Colorectal Cancer Patients Cohort; CRCSC, Colorectal Cancer Screening Cohort; NCRCC, National Colorectal Cancer Cohort; WHR, waist-to-hip ratio;
The unique strength of the NCRCC study includes many aspects: prospective and observational design, large population with geographic diversity, richness in epidemiological and clinical data, multiple biospecimens and follow-up strategies. The goal of the NCRCC study is using these strengths to specifically assess risk factors and biomarkers related to endpoints across the CRC continuum from the aetiology through survival.

The NCRCC study design allows for the comprehensive assessment of multiple factors related to CRC incidence, treatment, prognosis and survival, which has rarely been studied among Chinese population. Moreover, to the best of our knowledge, this large-scale national study (CRCPC), which includes 16 hospitals that are widely distributed in eastern, western, northern, southern and central China and has patients from almost all part of the country, is the largest CRC patient cohort with largest number of collaborating centres in China. Therefore, the results from the CRCPC may be generalisable to the whole patients with CRC in China.

The study is also strengthened by access to the EMR system to abstract multidimensional data (clinical, laboratory and imaging data) as well as the collection of comprehensive epidemiology data (dietary intake, lifestyle factors and medication use) at enrolment, which has not been available for most existing CRC patient cohort studies. Meanwhile, some exposure factors that are specific to the Chinese population offer a unique opportunity to better understand aetiology and prognosis of CRC. In addition, blood and tissue samples were collected from most participants, providing valuable resources to assess molecular biomarkers by multomics. A series of blood and urine samples in posttreatment periods were also collected, enabling the assessment of changes in biomarkers in prognosis. Of note, the CRCPC is one of few existing CRC patient cohorts with a large number of biospecimens.

However, there are several limitations to this study. First, the CRCPC is a multicentre study that may introduce heterogeneity in participant enrolment, sample collection and follow-up. However, several standardised protocols for the study have been established and strictly enforced at each research site. All the data are standardised and uploaded through the internet from each research site. Second, epidemiological data and specific biospecimen collection are not sufficient. In our cohort, some epidemiological factors (eg, lifestyle factors, dietary intake and dietary supplement use) and samples (eg, urine) are not available in all research sites. However, these limitations will be eliminated by our ongoing efforts to collect the above-mentioned data from the remaining participants of the cohort.

In conclusion, the NCRCC study provides a large database and biorepository to test a wide range of scientific hypotheses related to CRC in the era of precision prevention, prediction and treatment. In addition, investigation of the association between specific exposure patterns in the Chinese population and optimise evidence-based guidance for Chinese population. Together, future findings of the NCRCC study will provide research evidence from Chinese population and optimise evidence-based guidelines across the CRC continuum.

**Author affiliations**

1. Department of Colorectal Surgery and Oncology, Key Laboratory of Cancer Prevention and Intervention, Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
2. Department of Colorectal Surgery, Sun Yat-sen University Cancer Center, Guangzhou, China
3. Department of Colorectal Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
4. Department of General Surgery, Zhongshan Hospital Fudan University, Shanghai, China
5. Departments of Gastrointestinal Surgery and Pathology, West China Hospital, Sichuan University, Chengdu, China

**Table 4** Selected clinical characteristics of colorectal cancer patients in the CRCPC

| Selected clinical characteristics | CRCPC N=15 551 (%) |
|----------------------------------|--------------------|
| Tumour site                      |                    |
| Colon                            | 46.3               |
| Rectum                           | 52.3               |
| Synchronous                      | 1.4                |
| AJCC stage                       |                    |
| I                                | 16.4               |
| II                               | 34.9               |
| III                              | 34.5               |
| IV                               | 14.20              |
| Cell differentiation             |                    |
| Undifferentiation/poor differentiation | 18.4            |
| Moderate differentiation          | 9.2                |
| High differentiation             | 72.4               |
| Histology                        |                    |
| Adenocarcinoma                   | 94.7               |
| Mucinous adenocarcinoma           | 4.4                |
| Signet-ring cell carcinoma        | 0.5                |
| Carcinoid/neuroendocrine carcinoma | 0.3             |
| CEA                              |                    |
| <5 µg/L                          | 37.0               |
| ≥5 µg/L                          | 63.0               |
| CA199                            |                    |
| <37 µg/L                         | 9.4                |
| ≥37 µg/L                         | 90.6               |

AJCC, American Joint Committee on Cancer; CA199, carbohydrate antigen 199; CEA, carcinoembryonic antigen; CRCPC, Colorectal Cancer Patients Cohort.
Acknowledgements The authors sincerely acknowledge Prof. Zuo-Feng Zhang for his contributions to the study design. The authors would like to especially thank all NCRC participants as well as staffs in the following hospitals and research centres who have made substantial contributions to the establishment of the National Colorectal Cancer Cohort: Beijing Cancer Hospital (Beijing), Cancer Hospital, Chinese Academy of Medical Sciences (Beijing), Cancer Institute of Jianghan (Zhejiang), Changhai Hospital (Shanghai), Daping Hospital, Third Military Medical University (Chongqing), Fujian Provincial Cancer Hospital (Fuzhou, Fujian), Hubei Cancer Hospital (Wuhan, Hubei), Jiangsu Provincial People’s Hospital (Nanjing, Jiangsu), Peking University People’s Hospital (Beijing), Shaxi Provincial Cancer Hospital (Tayuan, Shanxi), Sun Yat-sen University Cancer Center (Guangzhou, Guangdong), The Second Affiliated Hospital, Zhejiang University School of Medicine (Hangzhou, Zhejiang), The Sixth Affiliated Hospital of Sun Yat-Sen University (Guangzhou, Guangdong), West China Hospital (Chengdu, Sichuan), Xijing Hospital, Fourth Military Medical University (Xi’an, Shaanxi), Zhejiang Academy of Medical Sciences (Hangzhou, China), Zhejiang Cancer Hospital (Hangzhou, Zhejiang), Zhejiang University Hospital Fudan University (Shanghai).

Collaborators The NCRC introduction, progress, publication and contact information can be found at http://www.ncrcrc.org.cn. Currently, the data are not yet openly available. Requests for potential collaboration can be addressed to Dr. Kefeng Ding (dingkefeng@zju.edu.cn).

Contributors YZ made substantial contributions to the conception and design of the study and drafted the manuscript. YH, XK and QM made contributions to the conception and design of the study and interpretation of data. ZZ, ZW, DW and JD made contributions to the design of the study and data collection. KD, TD, HC, HC, JZ and PC made contributions to the design of the study and interpretation of data. KD, TD made substantial contributions to the conception and design of the study. All authors have reviewed the manuscript critically for intellectual content and approved the final manuscript. KD acted as guarantor.

Funding This work was supported by the National Key R&D Program of China (2017YFC0909800).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Ethical approval for the NCRC study was granted by the Committe on Research involving Human Subjects of the Second Affiliated Hospital of Zhejiang University (2017-072). Written informed consent was provided by each participant.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data are available upon reasonable request. Interested researchers can submit project proposals and request data or samples by contacting dingkefeng@zju.edu.cn. Ethics approval will be necessary prior to obtaining data.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

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/.

ORCID iD
Kefeng Ding http://orcid.org/0000-0002-2380-3717

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;68:394–424.
2. Chen W, Zheng R, Baade P, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115–32.
3. Zhu J, Tan Z, Hollis-Hansen K, et al. Epidemiological trends in colorectal cancer in China: an ecological study. Dig Dis Sci 2017;62:235–43.
4. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin 2019;69:363–85.
5. Levin TR, Corley DA, Jensen CD, et al. Effects of organized colorectal cancer screening on cancer incidence and mortality in a large community-based population. Gastroenterology 2018;155:1383–91.
6. Lee JK, Jensen CD, Levin TR, et al. Long-Term risk of colorectal cancer and related deaths after a colonoscopy with normal findings. JAMA Intern Med 2019;179:153–60.
7. Ultrich CM, Gigie B, Böhm J, et al. The ColoCare study: a paradigm of Transdisciplinary science in colorectal cancer outcomes. Cancer Epidemiol Biomarkers Prev 2019;28:591–601.
8. Wu X, Hildebrandt MA, Ye Y, et al. Cohort profile: the MD Anderson cancer patients and survivors cohort (MDA-CPSC). Int J Epidemiol 2016;45:713–713f.
9. Liu P-H, Wu K, Ng K, et al. Association of obesity with risk of early-onset colorectal cancer among women. JAMA Oncol 2019;5:37–44.
10. Song M, Chan AT. The potential role of exercise and nutrition in harnessing the immune system to improve colorectal cancer survival. Gastroenterology 2018;155:596–600.
11. Song M, Chan AT, Diet CAT. Diet, gut microbiota, and colorectal cancer prevention: a review of potential mechanisms and promising targets for future research. Curr Colorectal Cancer Rep 2017;13:429–36.
12. Gong J, Hutter CM, Newcomb PA, et al. Genome-Wide interaction analyses between genetic variants and alcohol consumption and smoking for risk of colorectal cancer. PLoS Genet 2016;12:e1006296.
13. Rossi M, Jahanzaib Anwar M, Usman A. And alcohol Consumption-Populations to molecules. Cancers 2018;10:38.
14. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet 2019;51:76–87.
15. Johnson CM, Wei C, Ensor JE, et al. Meta-Analyses of colorectal cancer risk factors. Cancer Causes Control 2013;24:1207–22.
16. Song M, Chan AT, Zini J. Influence of the gut microbiome, diet, and environment on risk of colorectal cancer. Gastroenterology 2020;158:322–40.
17. Shah MS, DeSantis Tz, Weinmaier T, et al. Leveraging sequence-based faecal microbial community survey data to identify a composite biomarker for colorectal cancer. Gut 2018;67:882–91.
18. Liang QJ, Li T, Nakatsu G, et al. A novel faecal Lachnoclostridium marker for the non-invasive diagnosis of colorectal adenoma and cancer. Gut 2020;69:1248–57.
19. Shin DW, Choi YJ, Kim HS, et al. Secondary breast, ovarian, and uterine cancers after colorectal cancer: a nationwide population-based cohort study in Korea. Dis Colon Rectum 2018;61:1250–7.
20. Ye D, Huang Q, Li Q, et al. Comparative evaluation of preliminary screening methods for colorectal cancer in a mass program. Dig Dis Sci 2017;62:2532–41.
21. Fayers P, Bottomley A, et al., EORTC Quality of Life Group. Quality of life research within the EORTC-the EORTC QLQ-C30. European organisation for research and treatment of cancer. Eur J Cancer 2002;38 Suppl 4:S125–33.
22. Bray SE, Paulin FEM, Feng SC, et al. Gene expression in colorectal neoplasia: modifications induced by tissue ischaemic time and tissue handling protocol. Histopathology 2010;56:240–50.
23. World Health Organization. International guide for monitoring alcohol consumption and related harm. Geneva, Switzerland: World Health Organization, 2000.
24. Xiao Q, Lu W, Kong X, et al. Alterations of circulating bacterial DNA in colorectal cancer and adenoma: a proof-of-concept study. Cancer Lett 2021;499:201–8.
Supplementary online content

Cohort profile: The National Colorectal Cancer Cohort (NCRCC) study in China

Table S1 Hospitals involved in CRCPC and their locations
Table S1 Hospitals involved in CRCPC and their locations.

| Hospital                                        | City      | Province  |
|------------------------------------------------|-----------|-----------|
| Beijing Cancer Hospital                         | Beijing   |           |
| Cancer Hospital, Chinese Academy of Medical Sciences | Beijing   |           |
| Shanghai Hospital                               | Shanghai  |           |
| Daping Hospital, Third Military Medical University | Chongqing |           |
| Fujian Provincial Cancer Hospital                | Fuzhou    | Fujian    |
| Hubei Cancer Hospital                           | Wuhan     | Hubei     |
| Jiangsu Provincial People's Hospital             | Nanjing   | Jiangsu   |
| Peking University People's Hospital              | Nanning   | Beijing   |
| Shanxi Provincial Cancer Hospital                | Taiyuan   | Shanxi    |
| Sun Yat-sen University Cancer Center             | Guangzhou | Guangdong |
| The Second Affiliated Hospital, Zhejiang University School of Medicine | Hangzhou | Zhejiang |
| The Sixth Affiliated Hospital of Sun Yat-Sen University | Guangzhou | Guangdong |
| West China Hospital                             | Chengdu   | Sichuan   |
| Xijing Hospital, Fourth Military Medical University | Xi'an    | Shaanxi   |
| Zhejiang Cancer Hospital                        | Hangzhou  | Zhejiang  |
| Zhongshan Hospital Fudan University              | Hangzhou  | Shanghai  |