BMJ Open Gastroenterology

Baseline characteristics of participants recruited into the COLO-COHORT study: a UK-wide resource for colorectal cancer prevention

Yashvee Dunneram ,[1,2] Eleanor Hackney,[2,3] Sarah Manning ,[2,3] Panayiotis Louca,[1] Christopher J Stewart,[4] Guy S Taylor,[2] Sara Koo ,[3] Gregory R Young,[4] Suparna Mitra,[5] James S Hampton,[3] Christina Dobson,[2] Laura J Neilson,[2,3] Caroline Addison,[6] John Whelpton,[7] the COLO-COHORT research team,[2,3] Linda Sharp,[2] Mark A Hull ,[5] Colin Rees[2,3]

To cite: Dunneram Y, Hackney E, Manning S, *et al*. Baseline characteristics of participants recruited into the COLO-COHORT study: a UK-wide resource for colorectal cancer prevention. *BMJ Open Gastroenterol* 2026;**13**:e002421. doi:10.1136/bmjgast-2026-002421

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjgast-2026-002421).

YD, EH and SM contributed equally.

LS, MAH and CR are joint senior authors.

Received 14 May 2026
Accepted 11 September 2026

Check for updates

© Author(s) (or their employer(s)) 2026. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to
Professor Colin Rees;
colin.rees@newcastle.ac.uk

ABSTRACT

Objective To describe the design, recruitment, and baseline characteristics of the first phase of COLO-COHORT and to establish this population as a resource for colorectal cancer (CRC) prevention research.

Methods COLO-COHORT is a prospective, multicentre observational study recruiting adults undergoing colonoscopy through screening and routine National Health Service referral pathways within a 32-site UK network. We present descriptive data from the first 4754 participants, recruited through 27 Phase 1 sites, including demographics, lifestyle factors, clinical characteristics, biospecimen availability and colonoscopy outcomes.

Results Participants had a median age of 64 years (IQR 57–70), and 54.3% were male. Overweight (38.0%) and obesity (30.5%) were common, and cardiometabolic comorbidities were prevalent (11.8% with type II diabetes; 39.4% with hypertension). Colorectal neoplasia was identified at colonoscopy in over half of participants (56.1%), with histologically characterised data available on polyp type, size, morphology and anatomical location. Core demographic and clinical variables were available for >95% of participants, with colonoscopy findings recorded for all. Linked blood and stool microbiome biospecimens were available for approximately 80% and 40% of participants, respectively.

Conclusion This paper establishes and characterises COLO-COHORT's Phase 1 study cohort as a well-phenotyped resource, representing a population of individuals undergoing colonoscopy via UK screening and routine clinical pathways. This cohort will support and underpin future collaborative clinical, epidemiological and translational research in CRC prevention.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Large, well-characterised cohorts integrating colonoscopy findings with detailed lifestyle, clinical and biological data are needed to advance colorectal cancer (CRC) prevention research. COLO-COHORT represents one of the first large UK-wide prospective colonoscopy cohorts integrating detailed clinical, lifestyle and biological data.

WHAT THIS STUDY ADDS

⇒ This study introduces COLO-COHORT, a large UK multicentre colonoscopy cohort with linked demographic, lifestyle, clinical, biospecimen and colonoscopy data. The paper demonstrates the scale, completeness and research potential of the cohort as a resource for CRC prevention research.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ COLO-COHORT provides a platform for collaborative clinical, epidemiological and translational research to improve CRC risk stratification, investigate disease pathways and support future prevention and screening strategies.

INTRODUCTION

Colorectal cancer (CRC) remains one of the most commonly diagnosed cancers globally and a significant cause of cancer-related mortality.[1] In the UK, around 44 000 cases are diagnosed every year, with CRC being diagnosed most commonly in people over 50 years of age.[2 3] Prevention and early detection are critical in reducing CRC incidence and mortality. Pre-cancerous colorectal neoplasia, such as adenomatous and serrated polyps[4]—provide an opportunity to identify individuals at increased risk of CRC and to intervene before malignancy develops. Characterising the factors associated with all stages of colorectal neoplasia is therefore essential for informing prevention strategies, refining risk stratification in clinical and public health settings and optimising the use of endoscopy resources.

BMJ Open Gastroenterol: first published as 10.1136/bmjgast-2026-002421 on 1 October 2026. Downloaded from http://bmjopengastro.bmj.com/ on October 13, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

The COLO-COHORT study was established to: (1) develop a multi-factorial risk prediction model for colorectal neoplasia, (2) investigate the role of the gut microbiome in colorectal neoplasia and (3) establish a consent-for-contact (C4C) platform for future research.[5] COLO-COHORT additionally provides a valuable resource for investigating a range of clinical and epidemiological questions in colorectal neoplasia, including the use of multiple faecal immunochemical tests (FITs) in symptomatic participants and investigating factors associated with sessile serrated lesions (SSLs). As one of its kind in the UK, and by collecting extensive data across multiple domains, the study also offers a unique opportunity to study individuals with and without colorectal neoplasia at colonoscopy, supporting future CRC prevention research.

This paper presents a descriptive overview of participants recruited during Phase 1 of COLO-COHORT, comprising the first 4754 individuals recruited to the study. This paper aims to document how the cohort was assembled, describe baseline participant characteristics and demonstrate the completeness and breadth of available data. By formally introducing this resource at an early recruitment milestone, this paper seeks to support transparency, facilitate future collaboration and provide a foundation for subsequent hypothesis-driven collaborative analyses as recruitment continues.

METHODS

Study design and setting

COLO-COHORT is an ongoing multi-centre observational study conducted across 32 hospital sites in the UK. Adults aged ≥30 years scheduled to undergo colonoscopy are identified and recruited through National Health Service (NHS) research sites. Patients referred for colonoscopy through the Bowel Cancer Screening Programme (BCSPs), as well as routine clinical referral pathways (henceforth, non-BCSP)—including symptomatic referral, family history assessment and post-polypectomy or post-cancer surveillance—are eligible to participate. Recruitment to Phase 1 of the study, reported here, took place between 1 January 2020 and 31 July 2024 and represented a predefined recruitment milestone.

COLO-COHORT's published protocol[5] specifies recruitment targets of 10 000 participants for observational Group A—including 6000 in A1 and 4000 in A2—and 10 000 participants for the Group B consent-for-contact platform (COLO-SPEED), with recruitment planned over a 6-year period (2019–2025). Participants recruited to Group A may also participate in Group B; these targets therefore do not necessarily represent 20 000 unique individuals.

COLO-COHORT was designed to allow prospective follow-up through linkage to routinely collected NHS datasets. Ethical approval and participant consent were obtained to enable linkage to national cancer registration, hospital episode statistics and mortality records. While the current analysis focuses on cross-sectional baseline data from Phase 1 recruitment, planned data linkage will allow longitudinal follow-up for CRC incidence, related clinical outcomes, resource utilisation and mortality.

The recruitment process is summarised in figure 1. Further information on the study is available in the published protocol.[5]

This study is reported in accordance with the STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) guidelines for cohort studies; the completed checklist is provided in the online supplemental material.

Study population and data collection

Participants completed baseline demographic and lifestyle assessments and had colonoscopy findings recorded. Depending on the clinical pathway and recruitment arm (see protocol paper),[5] some participants also (optionally) provided blood samples and preprocedural FIT samples. Availability of specific data types, therefore, varied between participants. For this descriptive analysis, all participants were considered a single cohort, and no comparisons were made based on consent category or recruitment subgroup.

Data were captured using a centralised REDCap database[6] and underwent routine regular quality assurance procedures coordinated by the study centre, including range and internal consistency checks.

Demographics and clinical characteristics

Demographic data were collected by site research nurses, typically via telephone on the day prior to colonoscopy; a small number of sites collected these data in person on the day of the procedure (figure 1). Information recorded included age, sex, ethnicity, educational attainment, employment status, medical history, medication use, family history of CRC and other cancers and smoking status. Smoking status was subsequently categorised using pack-year categories (never, <10, 10–19, 20–29 and ≥30 pack-years). Anthropometric measurements, including height, weight and waist circumference, were measured by trained staff on the day of colonoscopy using standardised procedures. Waist circumference was measured in a standing position at the midpoint between the lower rib and the iliac crest using a non-stretch tape measure. Body mass index (BMI) was then calculated as weight in kilograms divided by height in metres squared. An overview of the data collected, including data completeness by instrument/data category, is summarised in online supplemental table 1.

Lifestyle factors

Diet

Dietary intake was assessed using the EPIC Food Frequency Questionnaire (FFQ),[7] which captured habitual intake of 130 food items over the preceding 12 months. Responses were converted into daily food and nutrient intakes using

BMJ Open Gastroenterol: first published as 10.1136/bmjgast-2026-002421 on 1 October 2026. Downloaded from http://bmjopengastro.bmj.com/ on October 13, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

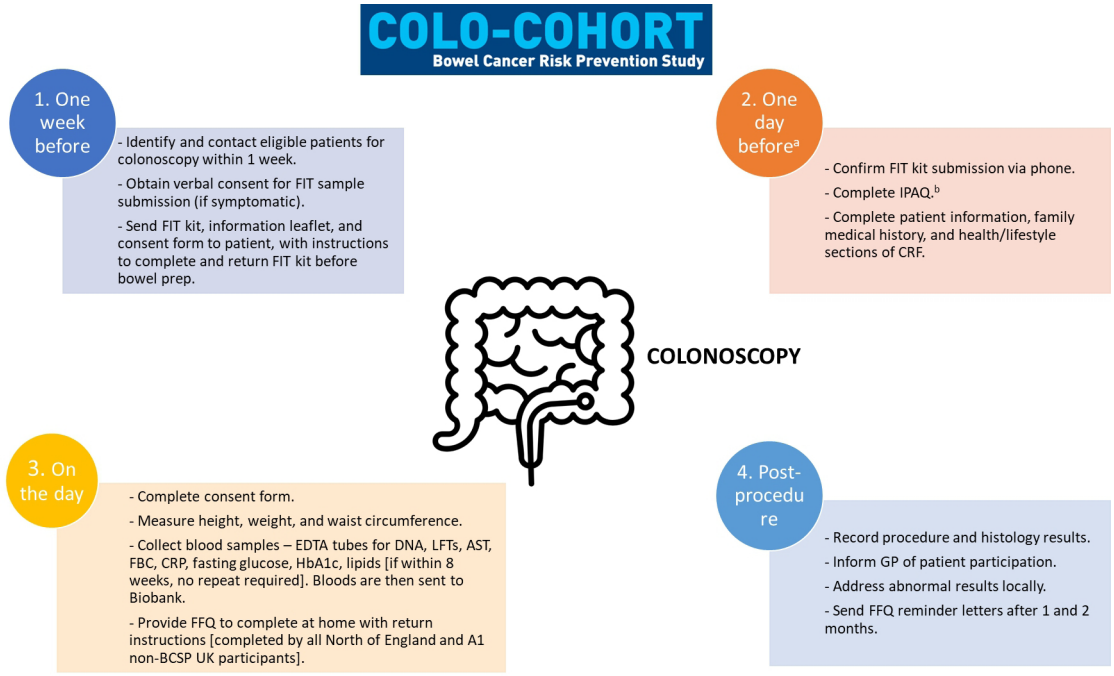


Figure 1 Participant recruitment and assessment process for Phase 1 of the COLO-COHORT study. [a]Except for two sites that completed the above on the day of the procedure. [b]Except for one site that provided the IPAQ to participants following colonoscopy to return with the FFQ. FFQ, Food Frequency Questionnaire; IPAQ, International Physical Activity Questionnaire.

the Food Extraction Tool for Analysis, based on the UK Nutrient Databank.[8]

Collection of dietary data was linked to the study's gut microbiome substudy so the FFQ was administered only to participants whose stool sample underwent microbiome analysis: this comprised all participants recruited through non-screening (symptomatic/other) pathways UK-wide, and participants recruited through the BCSP specifically in the North East of England, where screening FIT samples are additionally processed through the coordinating centre's microbiome pipeline.[5] BCSP participants recruited elsewhere had screening FIT samples processed through their local regional hub only and were therefore not included in the microbiome sub-study or sent an FFQ (figure 1).

Physical activity

Physical activity was assessed using validated questionnaires. Participants recruited at study outset completed the EPIC Physical Activity Questionnaire (n=270), while later participants completed the short form of the International Physical Activity Questionnaire (IPAQ)[9] (n=2616); this change was made to reduce participant burden. Physical activity data were expressed as metabolic equivalent task hours per week, and values derived from both instruments were standardised to this common metric to enable descriptive summarisation across data collection instruments.

FIT

Quantitative FIT samples were obtained pre-procedurally from participants, with participants instructed to collect samples prior to bowel preparation (according to consent and availability criteria). FIT concentrations were recorded as micrograms of haemoglobin per gram of faeces (µg Hb/g), with values outside analyser detection limits assigned the corresponding reporting threshold for descriptive purposes. Participants recruited through the BCSP had FIT samples taken as part of routine screening, while those recruited through other clinical pathways submitted a FIT sample at recruitment (research FIT), regardless of prior testing. In English symptomatic services, colonoscopy is typically offered to individuals with a FIT result ≥10 µg Hb/g, although thresholds vary across UK nations and clinical referral pathways. In English BCSP, colonoscopy is offered to individuals with a FIT result ≥120 µg Hb/g (currently being rolled out at 80 µg Hb/g); in Scotland, screening is offered at 80 µg Hb/g. FIT data were linked to clinical and colonoscopy findings for descriptive purposes. Where patients had previously submitted FIT samples for clinical analysis purposes, those FIT results were also recorded (where available).

Gut microbiome

Stool samples submitted for FIT testing were also used for gut microbiome analysis in participants who provided appropriate consent and from sites where collection could be performed. Microbiome sequencing data were available for a subset of participants, with 1762 individuals undergoing sequencing, as previously described,[10] with linked dietary and clinical data enabling future integrative analyses. Once the FIT sample was analysed,

BMJ Open Gastroenterol: first published as 10.1136/bmjgast-2026-002421 on 1 October 2026. Downloaded from http://bmjopengastro.bmj.com/ on October 13, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

pseudonymised samples were transferred to the Central Biobank Facility at Newcastle University and stored at −80 °C on arrival. Detailed laboratory and bioinformatic methods for microbiome sample processing and analysis are reported separately.[10] Gut microbiome data are available through the European Nucleotide Archive (ENA; www.ebi.ac.uk/ena/) under accession number PRJEB91658. Previous research by the team has demonstrated excellent agreement between microbiome profiles obtained from FIT samples and those from fresh stool analysis.[11 12]

Ascertainment of colorectal neoplasia

Colonoscopy findings were recorded for all participants, including the presence or absence of polyps and CRC. For polyps, information was collected on anatomical location, size, morphology, retrieval status and for retrieved polyps on histological subtype and dysplasia status. Neoplasia details were determined based on histological findings from hospital pathology reporting systems, with colorectal cancer site and stage recorded where applicable.

Neoplasia outcomes identified at colonoscopy were categorised as agreed a priori based on histological findings, with case definitions summarised in online supplemental table 2. These categories were developed following multidisciplinary discussion within the research team and informed by UK post-polypectomy and post-CRC surveillance guidelines.[13] Recognising that surveillance guidelines are influenced by both risk stratification and healthcare resource considerations at the patient level, neoplasia data were also coded flexibly to allow reclassification or regrouping of outcomes according to the specific aims of future analyses. A decision was made to include diminutive rectal hyperplastic polyps in the 'no neoplasia' group because of the lack of neoplastic potential, the current recommendation is not to remove them[13] and consequently, documentation and removal of these polyps are variable. This overall approach supports descriptive reporting and enables analyses at both the participant and lesion level.

Colonoscopy procedure details

Colonoscopy procedural details recorded included use of sedation and antispasmodic agents, withdrawal time (in minutes), performance of rectal retroflexion, completion of colonoscopy, quality of bowel preparation and an anonymised identifier for the endoscopist performing the procedure.

Blood samples

Blood samples were collected for routine clinical assays, including a full blood count, liver function tests, aspartate aminotransferase, C-reactive protein, lipid profile, fasting glucose and HbA1c testing. If relevant blood test results from the previous 8 weeks were available in the patient's records, repeat testing was not undertaken. A separate sample was taken for DNA extraction. All clinical blood tests—excluding those for DNA extraction—were labelled and processed according to local NHS Trust protocols in local laboratories. The sample for DNA was pseudonymised using the unique study ID and transferred to the Central Biobank Facility at Newcastle University. Samples were stored at −80°C in preparation for planned future DNA extraction.

Medications and supplement use

Information on medication use was collected at baseline as part of the clinical assessment. Participants were asked about use of selected medications relevant to CRC risk, including aspirin, statins, metformin and antibiotics. An open-ended field was also included to capture additional medications, from which use of proton pump inhibitors was identified. Information on calcium and vitamin D supplementation was also recorded.

Data quality assurance

Regular data checking and cleaning were conducted by the coordinating centre, with a different hospital site reviewed every 2 months. This entailed tracking down missing data, checking for implausible values, and following up with sites. In addition, approximately 5% of all paper Case Report Forms and FFQs were requested from hospital sites and cross-checked against REDCap entries to identify discrepancies and errors. This was carried out in two batches: the first ('Dataset 1') included participants who consented up to 30th June 2023 and showed a discrepancy rate of 0.9%; the second ('Dataset 2') included participants who consented between 1 July 2023 and 31 July 2024, with a discrepancy rate of 1.1%. The final dataset was exported to Stata (version 18) for further validation, outlier assessment and variable coding prior to analysis.

Statistical analysis

Participant characteristics were summarised using descriptive statistics, with the median and IQR reported for continuous variables and frequencies (with percentage values) for categorical variables. Descriptive summaries are presented overall and, where relevant, by recruitment pathway. Dietary intake, physical activity, anthropometric measures, smoking status and cardiometabolic comorbidities were summarised across colorectal lesion categories

For dietary variables, analyses were restricted to participants recorded as having returned an FFQ. Dietary data met the predefined inclusion criteria when no more than 10 FFQ food items were missing and total energy intake was plausible (women: 500–3500 kcal/day; men: 800–4000 kcal/day). Of 1802 participants recorded as having returned an FFQ, 1674 (92.9%) met these criteria and were included in dietary summaries. Dietary intake values were adjusted for total energy intake to account for potential misreporting.[14]

Patterns of missing data were examined descriptively, and analyses were conducted using available data for each variable, with exclusions applied only where specified.

BMJ Open Gastroenterol: first published as 10.1136/bmjgast-2026-002421 on 1 October 2026. Downloaded from http://bmjopengastro.bmj.com/ on October 13, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

RESULTS

Participant characteristics

Between 1 January 2020 and 31 July 2024, 4764 participants were recruited into Phase 1 of the COLO-COHORT study. Ten participants subsequently withdrew, leaving 4754 participants included in this analysis. Of these, 49% were recruited through the BCSP and 51% through symptomatic or other referral pathways.

Smoking exposure varied across pack-year categories: 45.1% of participants were never smokers (0 pack-years), 17.7% were light smokers (<10 pack-years), 11.7% were moderate smokers (10–19 pack-years), 8.9% were heavy smokers (20–29 pack-years) and 15.5% were very heavy smokers (≥30 pack-years) (table 1).

Participant characteristics varied by referral pathway. Individuals recruited through BCSP routes were slightly older on average, more likely to be male, and less likely to report current smoking or obesity compared with those recruited through non-screening pathways. Detailed characteristics by recruitment pathway are presented in table 1.

Phase 1 participants were recruited from 27 hospital sites: 26 in England (n=4636) and one in Scotland (n=118), within the wider 32-site COLO-COHORT network. Demographic and clinical characteristics varied across geographical areas of recruitment, including age distribution, sex, educational attainment, BMI and prevalence of cardiometabolic comorbidities. These differences illustrate the geographic and sociodemographic diversity of the cohort. Regional characteristics are summarised in online supplemental table 3, and Phase 1 recruitment numbers by site are presented in online supplemental table 4.

FIT measurements

FIT results were available for the majority of participants (97.2%), defined as having at least one recorded FIT result, including routine screening FIT results for BCSP participants and research FIT results for those recruited through non-screening pathways. Among non-BCSP participants with research FIT measurements, 1670 (72.3%) had values below 10 µg Hb/g, the threshold used to guide symptomatic referral in England. These research measurements should not be interpreted as the clinical FIT results used to determine referral. Screening programme FIT levels were above the applicable programme threshold of 120 µg Hb/g in England or 80 µg Hb/g in Scotland. The distribution of research FIT values is shown in online supplemental figure 1, illustrating the number of values and range of quantitative FIT data available for future analyses.

Blood biochemistry

Blood biochemistry results were available for 97% of participants, with data completeness summaries presented by recruitment pathway in online supplemental table 1. In addition, a total of 10 532 whole blood samples were collected and stored for future genomic analyses.

Colonoscopy findings and colorectal neoplasia

Colonoscopy findings were recorded for all participants. A total of 8149 polyps were identified across the cohort, of which 8036 (98.6%) had available histological confirmation. Polyp characteristics, including histological subtype, size, morphology, anatomical location and dysplasia status, are summarised in table 2.

At participant level, 2666 (56.1%) of 4754 participants had colorectal neoplasia identified at colonoscopy (table 3). This included 2117 participants (44.5%) with at least one adenoma, 888 (41.9%) with at least one advanced adenoma and 250 (5.3%) with CRC. At least one serrated lesion (defined as hyperplastic polyps, SSLs and traditional serrated adenomas) was identified in 923 participants (19.4%), including 193 (20.9%) with at least one advanced serrated lesion (serrated polyp >10 mm or serrated polyp with dysplasia) and 402 (8.5%) with one or more SSLs.

A total of 579 participants (12.2%) had both adenomatous and serrated lesions, 259 (5.4%) had both adenoma and SSL, and smaller proportions had cancer in combination with adenoma or serrated lesions (table 3).

Lesion distribution by demographic and clinical characteristics

Colorectal lesion distributions varied by demographic and clinical characteristics. Adenomas were more frequently observed in men, who accounted for 65.2% of participants with adenomas. The majority of colorectal lesions occurred in participants aged over 50 years, comprising more than 90% of cases across lesion categories (table 4).

Lifestyle and clinical data across lesion categories

Overall distributions varied across lesion groups, reflecting the breadth of lifestyle data available within the study population. For illustrative purposes, median processed meat intake ranged from approximately 19 to 23 g/day across lesion categories, compared with 17.4 g/day among participants with no neoplasia detected (table 4, online supplemental table 6). Detailed descriptive summaries of dietary intake, physical activity, anthropometric measures and smoking status are provided in online supplemental table 1 to characterise the range and completeness of available data.

Anatomical distribution varied by lesion type, with serrated lesions more commonly located in the proximal colon (66.3%) and advanced adenomas more frequently detected in the distal colon (35.9%). Lesion prevalence also differed by referral pathway, with advanced neoplasia more commonly identified among participants recruited through the BCSPs than through non-screening routes (table 4, online supplemental table 6).

BMJ Open Gastroenterol: first published as 10.1136/bmjgast-2026-002421 on 1 October 2026. Downloaded from http://bmjopengastro.bmj.com/ on October 13, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 1 Baseline demographic and clinical characteristics of Phase 1 participants

Risk factor	Total	BCSP	Non-BCSP
Age, years	64 (57–70)	64 (60–70)	62 (52–70)
Sex			
Female, n (%)	2171 (45.7)	869 (37.6)	1302 (53.3)
Male, n (%)	2582 (54.3)	1443 (62.4)	1139 (46.6)
Education level, n (%)			
Less than secondary/high school	1897 (39.9)	841 (36.4)	1056 (43.2)
Secondary/high school	969 (20.4)	437 (18.9)	532 (21.8)
University degree or equivalent professional qualification	1875 (39.4)	1024 (44.3)	851 (34.8)
Height, cm	1.70 (1.62–1.78)	1.73 (1.65–1.79)	1.69 (1.61–1.77)
Weight, kg	82.0 (70.0–95.0)	83 (72.0–95.3)	80.0 (68.3–93.4)
BMI, kg/m^2	27.8 (24.8–31.8)	27.9 (25.0–31.6)	27.8 (24.5–32.1)
BMI classification, n (%)			
Underweight (<18.5 kg/m^2)	54 (1.14)	21 (0.91)	33 (1.35)
Normal weight (18.5–24.9 kg/m^2)	1210 (25.5)	546 (23.6)	664 (27.2)
Overweight (25–29.9 kg/m^2)	1805 (38)	944 (40.8)	861 (35.3)
Obese (≥30–39.9 kg/m2)	1451 (30.5)	690 (29.8)	761 (31.2)
Severe obesity (≥40 kg/m^2)	196 (4.12)	91 (3.94)	105 (4.3)
Waist circumference, cm	99.5 (90-109)	101 (92–110)	98 (88–108)
Self-reported health conditions			
History of diabetes, n (%)			
Type I	38 (0.8)	22 (0.95)	16 (0.66)
Type II	561 (11.8)	237 (10.3)	324 (13.3)
No	4155 (87.4)	2053 (88.8)	2102 (86.1)
History of hypertension, n (%)			
Yes	1872 (39.4)	997 (43.1)	875 (35.8)
No	2879 (60.6)	1313 (56.8)	1566 (64.1)
History of bowel polyps, n (%)			
Yes	904 (19)	293 (12.7)	611 (25)
No	3844 (80.9)	2018 (87.3)	1826 (74.8)
History of colorectal cancer, n (%)			
Yes	178 (3.74)	15 (0.65)	163 (6.68)
No	4576 (96.3)	2297 (99.4)	2279 (93.3)
History of other cancers, n (%)			
Yes	632 (13.3)	313 (13.5)	319 (13.1)
No	4116 (86.6)	1996 (86.3)	2120 (86.8)
History of angina, n (%)			
Yes	417 (8.77)	190 (8.22)	227 (9.3)
No	4332 (91.1)	2121 (91.7)	2211 (90.5)
Medication history			
Aspirin, n (%)	519 (10.9)	250 (10.8)	269 (11.0)
Statin, n (%)	1809 (38.1)	929 (40.2)	880 (36.0)
Metformin, n (%)	402 (8.5)	176 (7.6)	226 (9.3)
Antibiotics in last 6 months, n (%)	909 (19.1)	351 (15.2)	558 (22.9)
Proton-pump inhibitor, n (%)	1773 (37.3)	707 (30.6)	1066 (43.7)
Lifestyle factors			

Continued

BMJ Open Gastroenterol: first published as 10.1136/bmjgast-2026-002421 on 1 October 2026. Downloaded from http://bmjopengastro.bmj.com/ on October 13, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 1 Continued

Risk factor	Total	BCSP	Non-BCSP
Calcium/vitamin D supplement use, n (%)			
Yes	1395 (29.3)	652 (28.2)	743 (30.4)
No	3349 (70.4)	1654 (71.5)	1695 (69.4)
Physical activity, tertiles n (%)			
Low	979 (20.6)	384 (16.6)	595 (24.4)
Medium	970 (20.4)	497 (21.5)	473 (19.4)
High	937 (19.7)	506 (21.9)	431 (17.6)
Smoking, n (%)			
Never smoker (0 pack-years)	2143 (45.1)	1035 (44.8)	1108 (45.4)
Light smoker (<10 pack-years)	841 (17.7)	393 (17.0)	448 (18.3)
Moderate smoker (10–19 pack-years)	556 (11.7)	252 (10.9)	304 (12.4)
Heavy smoker (20–29 pack-years)	425 (8.9)	213 (9.2)	212 (8.7)
Very heavy smoker (≥30 *pack*-years)	739 (15.5)	392 (17.0)	347 (14.2)
Alcohol, g/d	3.4 (0–10.2)	7.0 (1.6–12.5)	3.3 (0–9.9)

*Median and IQR or n (%); Percentages are calculated using the full column denominator. Missing or unknown values are not shown; consequently, percentages may not sum to 100%.
BCSP, Bowel Cancer Screening Programme.

DISCUSSION

This paper describes the baseline characteristics of the first 4754 participants recruited to Phase 1 of the COLO-COHORT study, a large UK-wide, multi-centre observational cohort designed to support CRC prevention research. By characterising participants recruited through both screening and other routine clinical pathways (mainly symptomatic referral) and by documenting the availability and completeness of lifestyle, clinical, procedural and biological data, this analysis establishes COLO-COHORT as a well-phenotyped data resource to support future clinical, epidemiological and translational research.

A key strength of COLO-COHORT is the availability of histologically confirmed colonoscopy outcomes, with detailed, comprehensive information on polyp type, size, morphology, dysplasia status and anatomical location. This level of clinical detail enables flexible categorisation of neoplasia outcomes and supports future analyses focused on specific lesion subtypes, including advanced adenomas and SSLs. Notably, the number of serrated lesions identified in COLO-COHORT represents one of the largest series reported within a colonoscopy research cohort to date. Given the growing recognition of the malignant potential of serrated lesions and their contribution to colorectal carcinogenesis, such scale and phenotypic detail are particularly valuable.[15] Despite increasing awareness, the aetiology and modifiable risk factors associated with serrated lesions remain incompletely understood, highlighting the importance of a well-phenotyped cohort such as COLO-COHORT for advancing colorectal prevention research.[16 17]

In this study population, more than half of the participants undergoing colonoscopy were found to have colorectal neoplasia, including adenomas, serrated lesions and CRC. The prevalence and anatomical distribution of lesions observed are consistent with patterns reported in colonoscopy-based populations, including a higher frequency of adenomas in men and a predominance of serrated lesions in the proximal colon.[16–21] These findings provide reassurance regarding the completeness and validity of lesion-level ascertainment and reflect the heterogeneous indications for colonoscopy across screening and non-screening referral pathways, including the increasing recognition and detection of serrated lesions as part of contemporary CRC prevention practice.

Participants in COLO-COHORT exhibited a high prevalence of established CRC risk factors, including overweight and obesity, cardiometabolic comorbidities and prior smoking. Lifestyle and behavioural data were comprehensively captured, including dietary intake, physical activity, anthropometric measures and smoking history. While this analysis was not designed to assess associations between lifestyle factors and colorectal neoplasia, the descriptive summaries presented illustrate the breadth and depth of exposure data available for future hypothesis-driven analyses. The inclusion of validated instruments for diet and physical activity assessment aligns with approaches commonly used in nutritional and lifestyle epidemiology and supports comparability with existing literature.[9 22 23] The distribution of overweight and obesity, cardiometabolic comorbidities and smoking history observed in COLO-COHORT is consistent with

BMJ Open Gastroenterol: first published as 10.1136/bmjgast-2026-002421 on 1 October 2026. Downloaded from http://bmjopengastro.bmj.com/ on October 13, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 2 Characteristics of polyps detected at colonoscopy among Phase 1 participants (n=4754)

Polyp characteristic	N (%)*
Total number of polyps	8036
Histological type	8036
Tubular adenoma	4417 (55.0)
Hyperplastic	1200 (14.9)
Tubulovillous adenoma	721 (8.97)
Sessile serrated lesion	659 (8.20)
Non-polypoid tissue	252 (3.14)
Inflammatory	108 (1.34)
Cancer	68 (0.846)
Villous adenoma	56 (0.697)
Other	48 (0.597)
Mixed polyp	38 (0.473)
Traditional serrated adenoma	15 (0.187)
Missing*	454 (5.65)
Size (mm)†	8036
<5	4246 (52.8)
5–9	2234 (27.8)
≥10	1452 (18.1)
Missing	104 (1.29)
Dysplasia present	8036
High-grade dysplasia	141 (1.75)
Low-grade dysplasia	5036 (62.7)
None	727 (9.05)
Missing/unknown	2132 (26.5)
Morphology	8036
Sessile (Is)	5036 (62.7)
Pedunculated (Ip)	918 (11.4)
Subpedunculated (Isp)	462 (5.75)
Flat elevated (0-IIa)	1334 (16.6)
Flat elevated and depressed (0-IIa/c)	29 (0.361)
Flat (0-IIb)	35 (0.436)
Flat depressed and elevated (0-IIc/a)	3 (0.037)
Depressed (0-IIc)	3 (0.037)
Missing	216 (2.69)
Location	8036
Proximal (3893)	
Caecum	886 (11.0)
Ascending colon	1344 (16.7)
Hepatic flexure	286 (3.56)
Transverse colon	1377 (17.1)
Distal (4135)	
Splenic flexure	312 (3.88)
Descending colon	666 (8.29)
Sigmoid	2206 (27.5)
Rectum	951 (11.8)
Missing	8 (0.10)

*Histology reports not updated or returned by hospitals.
†Based on the endoscopy report.

established CRC risk factor profiles reported in the UK and international literature and with patterns commonly observed among individuals undergoing colonoscopy and CRC screening.[23 24] Differences in lifestyle profiles between screening and non-screening populations have been reported previously and may reflect broader patterns of health-seeking behaviour and screening participation.[25] Together, these characteristics highlight the relevance of the population for future analyses examining the contribution of modifiable lifestyle factors to colorectal neoplasia risk.

COLO-COHORT also benefits from linkage to pre-procedural FIT results and from the availability of gut microbiome sequencing data in a substantial subset of participants. The use of FIT-derived stool samples for microbiome analysis has been previously validated,[11] enabling large-scale investigation of microbial composition and function in relation to colonoscopy-confirmed outcomes. These features position COLO-COHORT to address emerging research questions at the interface of lifestyle, host genomics, the gut microbiome and colorectal neoplasia, including the evaluation of strategies involving multiple FITs, an area of active investigation within CRC screening and symptomatic referral pathways.[26 27] In particular, the availability of detailed colonoscopy outcomes, FIT results and linked lifestyle and biological data provides an opportunity to build on existing work, such as COLO-FIT, a UK-based initiative evaluating FIT-based triage algorithms for symptomatic referral pathways,[28 29] and to explore whether additional factors may further inform risk stratification and referral algorithms used in clinical practice.

Few international cohorts combine colonoscopy-confirmed outcomes with detailed lifestyle and gut microbiome data at comparable scale. CRCbiome[30] in Norway is nested within a national FIT-based screening trial and combines lifestyle and gut microbiome data with screening outcomes, while the German BLITZ study[31] recruits participants undergoing screening colonoscopy and collects questionnaire, blood and stool data to investigate colorectal neoplasia risk and non-invasive detection markers. Both cohorts are restricted to screening populations, whose underlying risk profiles and routes to colonoscopy are likely to differ from those of COLO-COHORT participants recruited through non-BCSP clinical pathways. COLO-COHORT's inclusion of participants from both screening and non-BCSP pathways therefore provides a complementary resource spanning a broader range of colonoscopy indications.

Several limitations should be considered when interpreting the characteristics of the COLO-COHORT population. The cohort was predominantly White British, and a relatively high proportion of participants reported university-level education or equivalent qualifications. For example, recent census data indicate that approximately one-quarter to one-third of adults in comparable age groups in England and Scotland hold degree-level qualifications, suggesting that educational attainment in COLO-COHORT is higher

BMJ Open Gastroenterol: first published as 10.1136/bmjgast-2026-002421 on 1 October 2026. Downloaded from http://bmjopengastro.bmj.com/ on October 13, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 3 Participant-level colorectal neoplasia outcomes identified at colonoscopy among Phase 1 participants (n=4754)

		BCSP	Non-BCSP
Outcome category	n (%)*	n (%)*	n (%)*
No neoplasia	2088 (43.9)	591 (25.6)	1497 (61.3)
Low risk	1526 (32.1)	872 (37.7)	654 (26.8)
High risk	890 (18.7)	664 (28.7)	226 (9.3)
Cancer†	250 (5.3)	185 (8.0)	65 (2.7)
Any adenoma	2117 (44.5)	1451 (68.5)	666 (31.5)
Non-advanced adenoma	1229 (58.1)	786 (54.2)	443 (66.5)
Advanced adenoma	888 (41.9)	665 (45.8)	223 (33.5)
Any serrated lesion	923 (19.4)	609 (66.0)	314 (34.0)
Non-advanced serrated lesion	730 (79.1)	486 (79.8)	244 (77.7)
Advanced serrated lesion	193 (20.9)	123 (20.2)	70 (22.3)
Sessile serrated lesion (SSL)	402 (8.5)	296 (48.6)	106 (33.8)
Overlapping profiles			
Adenoma+SSL	259 (5.4)	208 (9.0)	51 (2.1)
Cancer+SSL	22 (0.5)	21 (0.9)	1 (0.1)
Cancer+adenoma	115 (2.4)	96 (4.2)	19 (0.6)
Cancer+serrated polyps	44 (0.9)	39 (1.7)	5 (0.2)
Serrated polyps+adenoma	579 (12.2%)	435 (18.8%)	144 (5.9%)
Any large non-pedunculated colorectal polyps	171 (3.6)	114 (4.9)	57 (2.3)

*Each outcome is reported independently. Participants may meet criteria for more than one category (eg, adenoma and SSL); therefore, totals exceed 100%.
†Out of the 250 cancers, 105 (42.0%) were found in the distal colon, 81 (32.4%) in the rectum and 70 (28.0%) in the proximal colon; six participants had cancers recorded at more than one anatomical location.
BCSP, Bowel Cancer Screening Programme.

than that of the general population.[32 33] While recruitment spanned multiple sites across England and Scotland, demographic characteristics vary regionally, and therefore the distributions observed in COLO-COHORT should be interpreted in the context of a colonoscopy-attending population drawn from specific geographic areas rather than as representative of the UK population overall. Available national datasets, including colonoscopy audit data and the National Endoscopy Database,[34 35] further indicate that individuals undergoing colonoscopy differ from the wider population with respect to sociodemographic characteristics. The patterns seen in COLO-COHORT may also reflect differential participation in colonoscopy-based studies and research more generally by socioeconomic and ethnic characteristics.[36 37] Limited ethnic diversity in study populations remains a recognised challenge in UK public health and clinical research and has been identified as a priority area for improvement by major funders.[38] These considerations should be borne in mind when interpreting future analyses derived from this cohort.

Beyond ethnic and educational composition, a more fundamental concern (which applies to many, if not most, research studies) relates to participation bias arising from the nature of consent to COLO-COHORT itself. Enrolment, for some, requires completion of a 130-item FFQ and a physical activity questionnaire, provision of stool and blood samples and consent to linkage with national NHS datasets—a substantially greater commitment than is required simply to attend a scheduled colonoscopy. Individuals who agree to this level of participation are likely to differ systematically from those who decline, not only in ethnicity and educational attainment but in unmeasured characteristics including health literacy, engagement with healthcare services and broader health-seeking behaviour. Because these characteristics may themselves be correlated with the lifestyle and clinical risk factors that COLO-COHORT is designed to study, the educational attainment gradient observed here (for example) is likely to be a partial marker of a broader participation effect rather than its full extent. This has direct implications for one of COLO-COHORT's stated aims—development of a multifactorial risk prediction model for colorectal neoplasia. Any such model derived from this cohort should be understood as describing risk relationships among people who both attend colonoscopy and consent to intensive research participation, and its generalisability to the wider colonoscopy-attending population, and more so to the general population, will require formal external validation before clinical application.

A related but distinct limitation concerns the population from which COLO-COHORT is drawn. By design, participants are recruited only after they have already been selected by the healthcare system for colonoscopy, either

BMJ Open Gastroenterol: first published as 10.1136/bmjgast-2026-002421 on 1 October 2026. Downloaded from http://bmjopengastro.bmj.com/ on October 13, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 4 Demographic, lifestyle, clinical and procedural characteristics according to colorectal lesion status

	Total	No neoplasia	Any adenoma	Non-advanced adenoma	Advanced adenoma	Any serrated lesion	Sessile serrated lesion only	Cancer only
Demographic characteristics								
Age								
<50 years	498 (10.5)	376 (18)	72 (3.4)	45 (3.66)	27 (3.04)	57 (6.18)	8 (9.3)	4 (4.35)
>50 years	4256 (89.5)	1712 (82)	2045 (96.6)	1184 (96.3)	861 (97)	866 (93.8)	78 (90.7)	88 (95.7)
Sex								
Female	2171 (45.7)	1163 (55.7)	737 (34.8)	468 (38.1)	269 (30.3)	357 (38.7)	44 (51.2)	47 (51.1)
Male	2582 (54.3)	924 (44.3)	1380 (65.2)	761 (61.9)	619 (69.7)	566 (61.3)	42 (48.8)	45 (48.9)
Lifestyle factors								
Diet								
Total energy intake, Kcal/d	1637 (1283–2046)	1656 (1294–2050)	1625 (1268–2034)	1627 (1263–2023)	1621 (1286–2036)	1612 (1224–1970)	1615 (1324–1805)	1729 (1261–2401)
Fruit intake, g/d	142.6 (61.9–252.3)	151.4 (66.9–254.7)	125.8 (55.6–244)	129.8 (61.9–242.7)	119.2 (47.3–245.2)	127.6 (51–241.3)	226.5 (100–339)	148.3 (62.6–231.9)
Vegetable intake, g/d	224.1 (139.5–311.9)	225.2 (139.5–314.4)	219 (143.2–306.9)	223.9 (146.7–314.3)	213.6 (139–302.1)	225.8 (137.1–309.4)	232.6 (160.8–288.6)	215.2 (121.8–321.5)
Red meat, g/d	20.3 (7.5–42.6)	19.3 (6.9–40.3)	24.2 (13.4–49)	24.8 (13.4–49)	22.2 (13.9–53.3)	21.7 (13.4–46)	19.3 (13.4–49)	17.1 (8.1–53.4)
Processed meat, g/d	19.6 (8.4–35.7)	17.4 (7.5–34.3)	21 (10.4–38.5)	20.7 (8.4–37.8)	23 (11.3–41.3)	20.7 (10.4–35.7)	22 (8.4–34.2)	18.7 (13.6–47.2)
Dietary calcium, mg/d	849.8 (647.3–1079.9)	858.9 (665.1–1104.9)	825.9 (629.4–1039.9)	834.6 (645.5–1054)	806.4 (614.1–1022.2)	804.2 (638.4–1042.4)	921.4 (658.1–1076.7)	937.9 (647.8–1116.3)
Alcohol intake, g/d	3.4 (0–10.2)	2.7 (0–9.1)	5.4 (0.5–11.9)	5 (0–11.2)	6 (1–13.6)	5 (0–10.3)	5 (0–10.3)	5.3 (0–21)
Physical activity								
Inactive	909 (34.7)	419 (36.2)	379 (32.8)	238 (35.6)	141 (29)	175 (33.2)	12 (25.5)	19 (38)
Moderately active	1205 (46.1)	536 (46.3)	529 (45.8)	283 (42.4)	246 (50.5)	241 (45.7)	23 (48.9)	24 (48)
Highly active	502 (19.2)	202 (17.5)	247 (21.4)	147 (22)	100 (20.5)	111 (21.1)	12 (25.5)	7 (14)
BMI								
Underweight (<18.5 kg/m^2)	54 (1.14)	31 (1.49)	18 (0.85)	11 (0.9)	7 (0.79)	5 (0.54)	0 (0.0)	0 (0.0)
Normal weight (18.5–24.9 kg/m^2)	1210 (25.5)	611 (29.3)	474 (22.4)	283 (23)	191 (21.5)	192 (20.8)	19 (22.1)	29 (31.5)
Overweight (25–29.9 kg/m^2)	1805 (38)	715 (34.2)	877 (41.4)	506 (41.2)	371 (41.8)	374 (40.5)	32 (37.2)	35 (38.0)
Obese (≥30–39.9 kg/m^2)	1451 (30.5)	622 (29.8)	652 (30.8)	378 (30.8)	274 (30.9)	303 (32.8)	31 (36)	27 (29.3)
Smoking								
Never smoker (0 pack-years)	2143 (45.1)	1035 (49.6)	873 (41.2)	524 (42.6)	349 (39.3)	336 (36.4)	34 (39.5)	43 (46.7)
Light smoker (<10 pack-years)	841 (17.7)	387 (18.5)	357 (16.9)	208 (16.9)	149 (16.8)	140 (15.2)	16 (18.6)	23 (25.0)
Moderate smoker (10–19 pack-years)	556 (11.7)	230 (11.0)	265 (12.5)	157 (12.8)	108 (12.2)	134 (14.5)	15 (17.4)	6 (6.5)
Heavy smoker (20–29 pack-years)	425 (8.9)	162 (7.8)	200 (9.4)	111 (9.0)	89 (10.0)	115 (12.5)	7 (8.1)	8 (8.7)
Very heavy smoker (≥30 pack-years)	739 (15.5)	254 (12.2)	398 (18.8)	214 (17.4)	184 (20.7)	190 (20.6)	12 (14.0)	12 (13.0)
Clinical characteristics								
History of diabetes								
No	4155 (87.4)	1840 (88.1)	1846 (87.2)	1072 (87.2)	774 (87.2)	799 (86.6)	76 (88.4)	75 (81.5)
Yes	599 (12.6)	248 (11.9)	271 (12.8)	157 (12.8)	114 (12.8)	124 (13.4)	10 (11.6)	17 (18.5)

Continued

BMJ Open Gastroenterol: first published as 10.1136/bmjgast-2026-002421 on 1 October 2026. Downloaded from http://bmjopengastro.bmj.com/ on October 13, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 4 Continued

	Total	No neoplasia	Any adenoma	Non-advanced adenoma	Advanced adenoma	Any serrated lesion	Sessile serrated lesion only	Cancer only
History of hypertension								
No	2879 (60.6)	1378 (66)	1181 (55.8)	686 (55.8)	495 (55.7)	525 (56.9)	61 (70.9)	45 (48.9)
Yes	1872 (39.4)	708 (33.9)	936 (44.2)	543 (44.2)	393 (44.3)	398 (43.1)	25 (29.1)	47 (51.1)
Procedural/lesion characteristics								
Indication route								
BCSPs	2312 (48.6)	591 (28.3)	1451 (68.5)	786 (64)	665 (74.9)	609 (66)	51 (59.3)	55 (59.8)
Non-BCSP	2442 (51.4)	1497 (71.7)	666 (31.5)	443 (36)	223 (25.1)	314 (34)	35 (40.7)	37 (40.2)
Polyp location								
Proximal	679 (14.3)	N/A	467 (22.1)	360 (29.3)	107 (12.1)	188 (20.4)	57 (66.3)	3 (3.26)
Distal	988 (20.8)	N/A	686 (32.4)	367 (29.9)	319 (35.9)	236 (25.6)	24 (27.9)	15 (16.3)
Both	1058 (22.3)	N/A	963 (45.5)	501 (40.8)	462 (52)	499 (54.1)	5 (5.81)	0 (0.0)

*Median and IQR; percentages may not total 100% due to missing data, which are not shown in the table.
BCSP, Bowel Cancer Screening Programme; BMI, body mass index; N/A, not applicable.

through engaging with screening and having a positive test or clinical referral following symptoms. This excludes individuals who do not participate in screening, do not present with or report symptoms, or are not referred for investigation—groups that are arguably of greatest relevance to a population-level CRC prevention strategy, since they are the least likely to be reached by existing pathways. COLO-COHORT is therefore well placed to address questions that (for example) compare individuals with and without colorectal neoplasia among those who reach colonoscopy, for example, identifying clinical, lifestyle or biological features that distinguish adenoma-bearing from adenoma-free colonoscopy patients—but this is conceptually distinct from identifying risk factors for colorectal neoplasia at the population level. Researchers using COLO-COHORT data should therefore distinguish between these two classes of question: those concerning risk stratification and phenotyping within a colonoscopy-attending population, for which the cohort is well suited, and those concerning population-level prevention or screening uptake, which would require linkage to or comparison with non-attending or non-referred populations beyond the scope of the current resource.

A further consideration relates to the interpretation of research FIT results among participants recruited through non-BCSP pathways. Among those with a research FIT result, 1670 (72.3%) had values below 10 µg Hb/g. As described in the Methods, research FIT was collected at recruitment regardless of previous clinical FIT testing and was not necessarily the test used to determine referral. Furthermore, the non-BCSP group included symptomatic, family-history, post-polypectomy surveillance and post-cancer surveillance referrals, to which the 10 µg Hb/g symptomatic threshold does not apply uniformly. Low research FIT values are therefore expected and do not imply inappropriate referral. As 99.22% of non-BCSP participants were recruited after 1 June 2020, the predominance of values below 10 µg Hb/g cannot principally be attributed to recruitment before formalisation of FIT-based symptomatic triage. Future analyses should distinguish research FIT from clinical referral FIT and account for referral indication when comparing screening and non-screening pathways.

Furthermore, lifestyle exposures were assessed at baseline and relied on self-reported questionnaires, which are subject to errors in recall and, hence, misclassification.[39] However, the dietary and physical activity instruments used have demonstrated acceptable reproducibility and validity in population-based studies.[7 9] Physical activity data were collected using two validated instruments over the course of recruitment, and although harmonised for descriptive purposes, direct comparability between instruments is limited. Future analyses using physical activity as a variable of interest should treat instrument type as a potential source of measurement heterogeneity, for example, by including it as a covariate or conducting instrument-stratified sensitivity analyses, rather than assuming full equivalence between the two harmonised measures.

BMJ Open Gastroenterol: first published as 10.1136/bmjgast-2026-002421 on 1 October 2026. Downloaded from http://bmjopengastro.bmj.com/ on October 13, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Additionally, some lesion histology data were unavailable due to non-retrieval or deferred management; however, these instances were infrequent and reflect real-world clinical practice.

The availability of dietary data reflected both the protocol-defined scope of FFQ administration and subsequent questionnaire return and quality assessment. FFQs were linked to the microbiome sub-study and were therefore not intended for all Phase 1 participants. Among participants recorded as having returned an FFQ, 1674 of 1802 (92.9%) met the dietary analysis criteria; this included 193 of 215 BCSP participants (89.8%) and 1481 of 1587 non-BCSP participants (93.3%). Thus, only 128 returned FFQs (7.1%) were excluded for missing items or implausible energy intake. Researchers planning nutritional epidemiology analyses should report the dietary analytic subsample clearly and consider its concentration in the North East of England and in non-screening participants, which may introduce selection effects into diet-focused analyses.

The principal sources of potential bias, their likely effects on different research objectives, proposed mitigation strategies and opportunities for external validation are summarised in online supplemental table 5.

Despite these limitations, COLO-COHORT represents a uniquely comprehensive nationwide resource for CRC prevention research. Ongoing and planned analyses will focus on risk prediction modelling, microbiome–neoplasia relationships, serrated lesion biology and longitudinal outcomes, building on the descriptive characterisation presented here as recruitment and follow-up progress. More broadly, the national scale of COLO-COHORT, inclusion of participants from both screening and routine clinical pathways, high-quality colonoscopy outcomes and extensive lifestyle and biological data provide a strong foundation for future collaborative clinical, epidemiological and translational research to advance CRC prevention.

Author affiliations
[1]Human Nutrition & Exercise Research Centre, Centre for Healthier Lives, Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK
[2]Population Health Sciences Institute, Newcastle University Centre for Cancer, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK
[3]Department of Gastroenterology, South Tyneside and Sunderland NHS Foundation Trust, South Tyneside District Hospital, South Shields, UK
[4]Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK
[5]Leeds Institute of Medical Research, University of Leeds, Leeds, UK
[6]North East Hub, Bowel Cancer Screening Programme, Gateshead, UK
[7]Patient and Public Involvement Lead, South Shields, UK

Acknowledgements We wish to acknowledge the Research and Innovation team at South Tyneside and Sunderland NHS Foundation Trust for helping to deliver the study as the sponsor site. Particular thanks to Head of Research Claire Livingstone, Lead COLOSPEED Nurse Ingrid Emmerson and research administrator Amy Burns. We also wish to acknowledge the assistance and laboratory expertise from the Newcastle Central Biobank Facility. We would like to extend our thanks to all participating patients and to the research teams at all participating sites for their dedication and hard work on this study: South Tyneside District Hospital, Sunderland Royal Hospital, North Tees University Hospital, Royal Victoria Infirmary, North Tyneside District Hospital, University Hospital of North Durham, South Tees Hospitals NHS Foundation Trust, Kettering General Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, West Hertfordshire Hospitals, Fairfield General Hospital, Gloucestershire Royal Hospital, Royal Bolton Hospital, Manchester Royal Infirmary, Wythenshawe Hospital, Royal Wolverhampton Hospital, Leeds General Infirmary, Manchester University, Salford Royal Hospital, Darlington Memorial Hospital, Hull Royal Infirmary, East Surrey Hospital, Golden Jubilee National Hospital, Cumberland Infirmary, West Cumberland Hospital, Stepping Hill Hospital, St George's Hospital, Maidstone Hospital, Tunbridge Wells Hospital, The Royal London Hospital, Mile End Hospital, Scunthorpe General Hospital, Diana, Princess of Wales Hospital, Colchester Hospital, Ipswich Hospital and St Bartholemew's Hospital.

Collaborators The COLO-COHORT research team was comprised of Richard Gallon, Ingrid Emmerson, Claire Livingstone, Amy Burns, Mark Graham, Nathania Bestwick, Amy Peasland, Aamar Ghouri, John Whelpton, Ajay Verma, Katharina Wallis, Sajal Rai, Tom Lee, Matthew Johnston, Sorena Afshar, John Greenaway, John Louis-Auguste, Andrew Day, Mark Canoy, Abby Conlin, Simon Hellier, Ian Farrell, Rosin Bevan, Susan Moug, Joe Geraghty, Noo Mohammed, Chris Mountford, Roger Prudham, Matthew Tutton, Ali Malik, Mo Thoufeeq, Shameer Mehta, S Sebastian and Kathy Dent.

Contributors YD: methodology; data curation; formal analysis; writing—original draft. EH: methodology; data curation; formal analysis; writing—original draft. SM: data curation. PL: data curation; writing—review and editing. CJS: conceptualisation; resources; writing—review and editing. GST: data curation; writing—review and editing. SK: writing—review and editing. GRY: resources; writing—review and editing. SM: resources; writing—review and editing. JSH: writing—review and editing. CD: writing—review and editing. LJN: writing—review and editing. CA: writing—review and editing. JW: contributed to the development and reviewed all patient-facing documents. LS: conceptualisation; methodology; writing, review and editing. MAH: conceptualisation; methodology; writing—review and editing. CR: conceptualisation; methodology; writing—review and editing. All authors reviewed and approved the final manuscript. CR is the guarantor and accepts full responsibility for the overall content of the work and the conduct of the study, had access to the research team and study records and controlled the decision to publish.

Funding This study is primarily funded by GUTS UK (Charity Registration No. 1137029, formerly known as Core), which administers an award by The Parabola Foundation, along with a funding award from The Sir Bobby Robson Foundation for Colo-Speed infrastructure and research delivery. CJS is funded by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (221745/Z/20/Z) and the 2021 Lister Institute Prize Fellow Award. G.R.Y. is supported by the NIHR Newcastle Biomedical Research Centre. LS is supported by the NIHR Newcastle Patient Safety Research Collaboration [NIHR204291]. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Competing interests LJN has received an unrestricted project grant from Medtronic and conference attendance support from Dr Falk Pharma and Tillotts. LS and CR have project grants from Medtronic and 3D-Matrix. All other authors declare no competing interests.

Patient consent for publication Not applicable.

Ethics approval This study involved human participants. The COLO-COHORT study was reviewed and approved by the West Midlands Edgbaston Research Ethics Committee (ref: 19/WM/0193, June 2019). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer-reviewed.

Data availability statement Data are available upon reasonable request. COLO-COHORT is designed as a collaborative research resource. The team are currently undertaking a range of analyses and once these are complete, the team will aim to make data available. Researchers interested in accessing data through collaborative projects are encouraged to contact the study team via the corresponding author. Data access will be subject to appropriate governance approvals and data sharing agreements.

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

BMJ Open Gastroenterol: first published as 10.1136/bmjgast-2026-002421 on 1 October 2026. Downloaded from http://bmjopengastro.bmj.com/ on October 13, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

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: https://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Yashvee Dunneram https://orcid.org/0000-0002-1012-7350
Sarah Manning https://orcid.org/0000-0002-7105-8109
Sara Koo https://orcid.org/0000-0001-7383-3140
Mark A Hull https://orcid.org/0000-0001-7414-1576

REFERENCES

1 Morgan E, Arnold M, Gini A, *et al*. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut* 2023;72:338–44.
2 Bowel Cancer UK. Bowel cancer [online]. 2024. Available: https://www.bowelcanceruk.org.uk/about-bowel-cancer/bowel-cancer/ [Accessed 21 Jul 2025].
3 Cancer Research UK. Bowel cancer statistics [online]. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer#BowelCS2 [Accessed 17 Jul 2025].
4 Leslie A, Carey FA, Pratt NR, *et al*. The colorectal adenoma-carcinoma sequence. *Br J Surg* 2002;89:845–60.
5 Hampton JS, Koo S, Dobson C, *et al*. The COLO-COHORT (Colorectal Cancer Cohort) study: protocol for a multicentre, observational research study and development of a consent-for-contact research platform. *Colorectal Dis* 2022;24:1216–26.
6 Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.
7 Bingham SA, Welch AA, McTaggart A, *et al*. Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. *Public Health Nutr* 2001;4:847–58.
8 Public Health England. Composition of foods integrated dataset (CoFID) [Internet]. 2019. Available: https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid [Accessed 15 Mar 2024].
9 Craig CL, Marshall AL, Sjöström M, *et al*. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 2003;35:1381–95.
10 Louca P, Manning S, Hackney E, *et al*. Gut microbiome signatures in colorectal neoplasia: a cross-sectional study across neoplasia stages and subtypes. *Gut* 2026.:gutjnl-2025-337478.
11 Masi AC, Koo S, Lamb CA, *et al*. Using faecal immunochemical test (FIT) undertaken in a national screening programme for large-scale gut microbiota analysis. *Gut* 2021;70:429–31.
12 Mitra S, J Stewart C, Nelson A, *et al*. Long-term stability of the faecal microbiome profile in faecal immunochemical test (FIT) samples. *Gut* 2025;74:e8.
13 Rutter MD, East J, Rees CJ, *et al*. British Society of Gastroenterology/Association of Coloproctology of Great Britain and Ireland/Public Health England post-polypectomy and post-colorectal cancer resection surveillance guidelines. *Gut* 2020;69:201–23.
14 Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr* 1997;65:1220S–1228S.
15 Mallick TR, Hasan M. Serrated polyps in the colorectum: clinical significance and future directions. *BMJ Open Gastroenterol* 2025;12:e001927.
16 Wong MCS, Huang J, Huang JLW, *et al*. Global prevalence of colorectal neoplasia: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2020;18:553–61.
17 Bretthauer M, Kaminski MF, Løberg M, *et al*. Population-based colonoscopy screening for colorectal cancer: a randomized clinical trial. *JAMA Intern Med* 2016;176:894–902.
18 Rex DK, Ahnen DJ, Baron JA, *et al*. Serrated lesions of the colorectum: review and recommendations from an expert panel. *Am J Gastroenterol* 2012;107:1315–29.
19 Corley DA, Jensen CD, Marks AR, *et al*. Variation of adenoma prevalence by age, sex, race, and colon location in a large population: implications for screening and quality programs. *Clin Gastroenterol Hepatol* 2013;11:172–80.
20 East JE, Vieth M, Rex DK. Serrated lesions in colorectal cancer screening: detection, resection, pathology and surveillance. *Gut* 2015;64:991–1000.
21 Bone E, Kumar S, Richards S, *et al*. Current sessile serrated lesion incidence: implications for future clinical practice. *ANZ J Surg* 2024;94:1806–11.
22 Mulligan AA, Luben RN, Bhaniani A, *et al*. A new tool for converting food frequency questionnaire data into nutrient and food group values: FETA research methods and availability. *BMJ Open* 2014;4:e004503.
23 World Cancer Research Fund/American Institute for Cancer Research. Continuous update project expert report: diet, nutrition, physical activity and colorectal cancer [online]. 2018. Available: https://www.wcrf.org/wp-content/uploads/2024/10/Colorectal-cancer-report.pdf [Accessed 10 Nov 2025].
24 Asgari A, Ziamanesh F, Aliasgari A, *et al*. Incidence and predicting factors of inadequate bowel preparation for colonoscopy: a cross-sectional study. *JGH Open* 2024;8:e13116.
25 von Wagner C, Baio G, Raine R, *et al*. Inequalities in participation in an organized national colorectal cancer screening programme: results from the first 2.6 million invitations in England. *Int J Epidemiol* 2011;40:712–8.
26 Lim TZ, Lau J, Wong GJ, *et al*. Colorectal cancer in patients with single versus double positive faecal immunochemical test results: a retrospective cohort study from a public tertiary hospital. *PLoS One* 2021;16:e0250460.
27 Gerrard AD, Maeda Y, Miller J, *et al*. Double faecal immunochemical testing in patients with symptoms suspicious of colorectal cancer. *Br J Surg* 2023;110:471–80.
28 Rees C, Hamilton W. Overview of the COLOFIT study. *Frontline Gastroenterol* 2025.
29 Crooks CJ, West J, Jones J, *et al*. COLOFIT: development and internal-external validation of models using age, sex, faecal immunochemical and blood tests to optimise diagnosis of colorectal cancer in symptomatic patients. *Aliment Pharmacol Ther* 2025;61:852–64.
30 Kværner AS, Birkeland E, Bucher-Johannessen C, *et al*. The CRCbiome study: a large prospective cohort study examining the role of lifestyle and the gut microbiome in colorectal cancer screening participants. *BMC Cancer* 2021;21:930.
31 Fu R, Chen X, Seum T, *et al*. Red and processed meat intake, polygenic risk and the prevalence of colorectal neoplasms: results from a screening colonoscopy population. *Nutrients* 2024;16:2609.
32 Office for National Statistics. *Education, England and Wales: Census 2021 [online]*. London, 2023. Available: https://www.ons.gov.uk/
33 National Records of Scotland. Scotland's census 2022: education, labour market and travel to work [online]. 2024. Available: https://www.scotlandscensus.gov.uk/ [Accessed 8 Feb 2026].
34 Beaton DR, Sharp L, Lu L, *et al*. Diagnostic yield from symptomatic gastroscopy in the UK: British Society of Gastroenterology analysis using data from the National Endoscopy Database. *Gut* 2024;73:1421–30.
35 Beaton D, Sharp L, Lu L, *et al*. Diagnostic yield from symptomatic lower gastrointestinal endoscopy in the UK: a British Society of Gastroenterology analysis using data from the National Endoscopy Database. *Aliment Pharmacol Ther* 2024;59:1589–603.
36 Routen A, Bodicoat D, Willis A, *et al*. Tackling the lack of diversity in health research. *Br J Gen Pract* 2022;72:444–7.
37 National Institute for Health and Care Research (NIHR). Randomised controlled trial participants diversity data report [online]. 2022. Available: https://www.nihr.ac.uk/randomised-controlled-trial-participants-diversity-data-report [Accessed 10 Feb 2026].
38 National Institute for Health and Care Research (NIHR). Improving inclusion of under-served groups in clinical research: guidance from the INCLUDE project [Internet]. 2024. Available: https://www.nihr.ac.uk/documents/improving-inclusion-of-under-served-groups-in-clinical-research-guidance-from-include-project/25435 [Accessed 10 Feb 2026].
39 Freedman LS, Commins JM, Moler JE, *et al*. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *Am J Epidemiol* 2014;180:172–88.

BMJ Open Gastroenterol: first published as 10.1136/bmjgast-2026-002421 on 1 October 2026. Downloaded from http://bmjopengastro.bmj.com/ on October 13, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.