BMJ Open

Fibre-rich Foods to Treat Obesity and Prevent Colon Cancer trial study protocol: a randomised clinical trial of fibre-rich legumes targeting the gut microbiome, metabolome and gut transit time of overweight and obese patients with a history of noncancerous adenomatous polyps

Terryl J Hartman [1,2] Jennifer Christie,[3] Annette Wilson [4] Thomas R Ziegler,[5] Barbara Methe,[6] William Dana Flanders,[7] Barbara J Rolls [8] Blaine Loye Eberhart [4] Jia V Li [9] Helaina Huneault [2] Ben Cousineau [2] Miriam R Perez [4] Stephen J D O'Keefe[4]

To cite: Hartman TJ, Christie J, Wilson A, *et al*. Fibre-rich Foods to Treat Obesity and Prevent Colon Cancer trial study protocol: a randomised clinical trial of fibre-rich legumes targeting the gut microbiome, metabolome and gut transit time of overweight and obese patients with a history of noncancerous adenomatous polyps. *BMJ Open* 2024;**14**:e081379. doi:10.1136/bmjopen-2023-081379

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2023-081379).

Received 25 October 2023
Accepted 15 January 2024

For numbered affiliations see end of article.

Correspondence to
Professor Terryl J Hartman; tjhartm@emory.edu

ABSTRACT

Introduction Recently published studies support the beneficial effects of consuming fibre-rich legumes, such as cooked dry beans, to improve metabolic health and reduce cancer risk. In participants with overweight/obesity and a history of colorectal polyps, the Fibre-rich Foods to Treat Obesity and Prevent Colon Cancer randomised clinical trial will test whether a high-fibre diet featuring legumes will simultaneously facilitate weight reduction and suppress colonic mucosal biomarkers of colorectal cancer (CRC).

Methods/design This study is designed to characterise changes in (1) body weight; (2) biomarkers of insulin resistance and systemic inflammation; (3) compositional and functional profiles of the faecal microbiome and metabolome; (4) mucosal biomarkers of CRC risk and (5) gut transit. Approximately 60 overweight or obese adults with a history of noncancerous adenomatous polyps within the previous 3 years will be recruited and randomised to one of two weight-loss diets. Following a 1-week run-in, participants in the intervention arm will receive preportioned high-fibre legume-rich entrées for two meals/day in months 1–3 and one meal/day in months 4–6. In the control arm, entrées will replace legumes with lean protein sources (eg, chicken). Both groups will receive in-person and written guidance to include nutritionally balanced sides with energy intake to lose 1–2 pounds per week.

Ethics and dissemination The National Institutes of Health fund this ongoing 5-year study through a National Cancer Institute grant (5R01CA245063) awarded to Emory University with a subaward to the University of Pittsburgh. The study protocol was approved by the Emory Institutional Review Board (IRB approval number: 00000563).

Trial registration number NCT04780477.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This study offers a comprehensive analysis of the effects of a high-fibre diet on human microbiota, metabolome and colonic mucosal biomarkers of colorectal cancer (CRC) over a 12-month period, while also evaluating the long-term benefits of nutrition education on reducing obesity and CRC risk at ~3 years.

⇒ Analysis of faecal short-chain fatty acids, indicative of fibre consumption and bile acids, representative of fat intake, aids in monitoring dietary adherence and determining if intensified behavioural changes and fibre supplementation are required.

⇒ Dietary compliance in the intervention group is a potential limitation, and the study is not sufficiently powered to evaluate if the anticipated changes in mucosal biomarkers predict polyp recurrence or malignant transformation.

⇒ Given that this cancer prevention study specifically targets healthy individuals, future results may not be generalisable to patients with cancer.

BACKGROUND

Colorectal cancer (CRC) is the third most common form of cancer in the USA.[1] Obesity increases the risk of at least 13 cancers, including CRC.[2 3] Burkitt's original hypothesis[4] from 1963 highlights that westernised diseases such as CRC and obesity may result from fibre deficiency from the commercial refinement

of foods. Many plausible mechanisms explain why high-fibre diets, especially a high-legume diet (HLD), may reduce CRC risk. First, fibre is fermented by the colonic microbiota to produce short-chain fatty acids (SCFAs). The SCFA butyrate has a remarkable array of colonic mucosal health-promoting, anti-inflammatory and antineoplastic properties.[5 6] Second, microbiota break down plant cell walls releasing phytochemicals, which also have powerful anti-inflammatory and anticarcinogenic effects.[7 8] Third, colonic transit is accelerated, reducing contact time with luminal carcinogens, such as heterocyclic amines formed from cooked red meat,[9] and secondary bile acids (BA), induced by a high-fat diet and synthesised by the colonic microbiota.[10 11]

A recent randomised controlled feeding study incorporated a 2-week food exchange, where African American subjects from Pittsburgh were fed a high-fibre (~50 g/day), low-fat African-style diet and rural Africans were fed a high-fat, low-fibre western-style diet. Results suggested that within weeks, mucosal and faecal biomarkers of cancer risk responded favourably to the high-fibre diet, with proliferative rates and inflammatory biomarkers decreasing and microbiota composition adapting to increase butyrogenesis.[12]

Our prior research suggests that fibre may reduce cancer risk indirectly by promoting weight loss, improving insulin sensitivity and decreasing inflammation.[5 13–15] On average, individuals consume a similar weight of food daily; thus, replacing energy-dense foods with lower energy density foods, such as legumes, should potentiate weight control.[16] Legumes are high in resistant starch, insoluble fibre and especially soluble fibre. Therefore, legumes absorb water during digestion, increasing viscosity, encouraging stomach distension and inducing satiation.[17]

Fibre-rich diets may also affect other physiological mechanisms important for weight control.[18 19] Trypsin inhibitors and other bioactive compounds found in legumes (eg, lectins) may directly stimulate cholecystokinin secretion in the proximal intestine to increase satiety.[20–22] SCFA production may play a role in appetite regulation through stimulation of anorexic gut hormones, peptide YY (PYY) and glucagon-like peptide (GLP-1), slowing gastric emptying.[20 23] Moreover, microbial acetate has been shown to suppress appetite through central hypothalamic mechanisms involving changes in transcellular neurotransmitter cycles.[24]

Emerging human evidence links the gut microbiome, insulin resistance, inflammation and obesity with adenomatous polyps and colon cancer.[23 25–28] This study will provide an opportunity to characterise gut motility, microbial changes and metabolome composition profiles that may influence weight loss and have a role in the prevention of adenomas and CRC, providing novel and potentially therapeutic information.

Study aims and outcome measures

The overall aim of this study is to perform a randomised controlled trial of an HLD compared with a control diet in 60 highest-risk middle-aged participants to measure its ability to reduce body weight. Biomarkers of insulin resistance, systemic inflammation, gut transit and colon cancer risk will be included. We hypothesise that restoring the diet with natural high-fibre content, principally with legumes, will lead to a more significant weight loss and improvements in biomarkers associated with colon cancer risk compared with a control diet. See table 1 for a detailed timeline of outcome measures.

METHODS AND ANALYSIS

This study is a parallel-arm randomised clinical trial in overweight/obese healthy persons with a history of noncancerous adenomatous polyp(s). Investigators will be blinded to the diet treatment; however, participants may be able to discern which diet they are randomised to. Preportioned entrées will be provided with regular nutrition education sessions with a dietitian.

Patient and public involvement

Patients and/or the public will not be involved in the design, conduct, reporting or dissemination plans of this research.

Participant recruitment

We aim to accrue 60 middle-aged adults (50% male, 50% female) using a combination of targeted advertisements in the Emory Gastroenterology (GI) Clinics and mailings sent to individuals who may be eligible because of their colonoscopy results.

Eligibility criteria

Inclusion

(1) Free-living adults 40–75 years old, (2) body mass index (BMI) 25–40 kg/m^2, (3) colonoscopy within 3 years that found/removed ≥1 adenoma >0.5 cm, (4) English speaking, (5) ambulatory and (6) able to provide informed consent.

Exclusion

(1) Serious medical condition, (2) history of CRC, bowel resection, polyposis syndrome or inflammatory bowel disease, (3) smoked regularly in the past year, (4) dietary restrictions substantially limiting compliance (5) planning on substantially changing usual exercise behaviour, (6) regular use of medication that may interfere with study procedures and (7) women currently pregnant, breast feeding or planning a pregnancy.

Informed consent

Eligible participants will be invited to an in-person screening at the Clinical and Translational Science Alliance at Emory University (CTSA). After signed informed consent is obtained, we will conduct the standard

Table 1 Intervention and assessment activities timeline

(W=weekly; M=monthly; B=bimonthly, X=once, N=if needed)														
Phase	**BL**	**Intervention (entrées, instruction)**						**Self-direction**			**Maintenance**			**Ext**
Month	0	1	2	3	4	5	6	7	8	9	10	11	12	~36
Food provision	W*	W	W	W	W	W	W							
Nutritionist (in person or Zoom)		W	W	W	W	W	W	M	M	M	M	M	M	N
Nutritionist (phone, email, text)		W	W	W	W	W	W	B	B	B	B	B	B	M
Assessments														
Mucosal biopsy	X						X						X	X
Stool sample for microbiome and metabolomics	X						X						X	X
Faecal fibre consumption biomarkers (SCFA/BA)	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Urine, first AM	X						X						X	X
Gastric emptying test	X						X							
Bowel function questionnaire	X						X						X	
Body weight	X	W	W	W	W	W	W	M	M	M	M	M	M	X
Waist circumference	X						X						X	
DXA	X						X						X	
Blood—fasting insulin (mIU/L), blood glucose (mg/dL), hs-CRP (mg/L)	X						X						X	X
Diet recalls (2/time point), run-in	X			X			X			X			X	X
Physical activity assessment	X						X						X	X
Pedometers	X						X						X	X
RISE-Q[59], DSQ[58]	X						X						X	
VAS[60]	X	X	X	X	X	X	X	X	X	X	X	X	X	
Adherence contacts		W	W	W	W	W	W	B	B	B	B	B	B	M

*A 1-week run-in diet of control diet entrées and sides at weight maintenance energy level is provided to all participants before baseline assessments and randomisation. Ext=extended follow-up (when a participant has a subsequent colonoscopy).
BA, bile acids; DSQ, Diet Satisfaction Questionnaire; DXA, dual-energy X-ray absorptiometry; hsCRP, high sensitive C reactive protein; RISE-Q, Reasons Individuals Stop Eating Questionnaire; SCFA, short-chain fatty acid; VAS, Visual Analogue Scale.

screening tests required for healthy participants. A separate consent will also be obtained for permission to store biological samples for future studies related to obesity and CRC.

Confidentiality

Confidentiality will be assured by using subject codes rather than personal identifiers. Any electronic data will be encrypted and accessible only with a login and protected password by the study staff. A certificate of confidentiality from the National Institutes of Health (NIH) has been attained. After the study is completed, all data and specimens will be kept secure according to NIH and Food and Drug Administration (FDA) regulations.

Study intervention

Run-in phase

Before randomisation, participants will proceed through a 1-week run-in where foods representative of the control diet will be provided. The run-in helps to standardise conditions and provides confidence in the participant's ability to adhere to the study protocol.

The study statistician will use the default random number generator in the R Software program V.4.1.3. to allocate subjects to each treatment arm. To conceal the randomisation sequence to eligible participants and study investigators, we will use numbered, opaque envelopes that contain the treatment assignment and allocate men and women separately for enrollment by the study coordinator.

Diet and nutrition education

We will use the Mifflin-St. Jeor equation to estimate energy needs for weight maintenance reducing this value to facilitate weight loss of 1–2 lbs/week (minus ~500–1000 kcal/day).[29]

Prepared portion control entrées

To all participants, we will provide preportioned entrées for two meals/day in months 1–3 and one meal/day in months 4–6. The HLD group will receive entrees from a menu cycle developed with a standard set of legumes primarily from the Phaseolus vulgaris species (eg, navy, pinto, black, kidney beans) to limit nutrient and phytochemical variability. The diet will contain approximately 250 g of legumes per day (~1 ½ cups cooked). This level will add approximately 30 g of dietary fibre/day from the legume dishes, ensuring a total intake of ~45–50 g/day. Previously, this level reduced colonic mucosal biomarkers of cancer risk within 2 weeks and is associated with minimal colon cancer risk in rural Africans[12 30] and a reduction in large polyp recurrences in the Polyp Prevention Trial.[31] The control group will also receive preportioned meal replacement entrées with legumes replaced by lean chicken/meat. Please see online supplemental materials 1, for example, meal plans for each group. All entrées will be prepared, preportioned and stored at the CTSA under the supervision of the bionutritionist. A printed sheet will be provided to record the amount of each entrée eaten. Education about the consumption of ad libitum sides tailored for weight loss was provided with the American Diabetes Associated food lists for weight management serving as a general guide.[32] The intake of sides contributing to total energy intake was not controlled by the study to enable evaluation of the role of legumes in promoting control of self-selected food and energy intake.

Self-direction and maintenance

Participants will continue on their respective diets in months 7–9 but will assume responsibility for food preparation. Long-term weight control is associated with frequent self-monitoring (eg, weight checks), replacement of high-energy-density foods with lower energy density alternatives and portion control, among other strategies.[33] Skill-building and behavioural strategies to address the aforementioned behaviours will be incorporated during bionutritionist encounters at the time points outlined in table 1. Participants in both arms received comparable nutrition advice at equal time points with a focus on weight management and action-oriented eating behaviour tips. During months 10–12, participants will interact with study staff monthly for follow-up and support.

Extended follow-up

Twelve months is adequate time to assess changes in mucosal biomarkers of CRC risk, but not for assessing polyp recurrence or cancer development. Consequently, to explore the long-term success of our diet behaviour modification training on weight control, mucosal biomarker suppression, polyp recurrence and carcinogenesis, we will extend the follow-up to the participants' next routine surveillance colonoscopy. This will be exploratory as the numbers will likely be insufficient to show significant reductions in polyps or cancer but will provide essential data for a definitive large-scale population study aimed at increasing the consumption of plant-based foods and reducing the risk of Westernised diseases should our intervention prove positive. We will ask participants to notify us of the scheduling of surveillance colonoscopies and request information on the size, multiplicity, anatomic location and histology of any polyps or cancer.

Data collection

Anthropometry

Bodyweight, height and waist circumference will be measured using the most recent National Health and Nutrition Examination Survey (NHANES) procedures.[34] We will also give participants smart scales, such as the Fitbit Aria (https://www.fitbit.com/global/us/products/scales/aria-air), for weekly home self-monitoring. Changes in body fat, composition and distribution will be assessed using dual-energy X-ray absorptiometry.

Inflammation/insulin sensitivity

Fasting blood samples will be collected, aliquoted and stored frozen at −80°C for analysis of biomarkers of insulin

resistance and systemic inflammation (see table 1). Please see online supplemental materials 2 for further details on these methods.

Stool and urine sample collection

Procedures developed in our NIH-supported studies will be used to ensure scientific rigour for collecting and analysing stool and urine samples. All samples will be aliquoted and held at –80°C for future DNA extraction and microbiome and metabolome analysis. Please see online supplemental materials 3 for further details on these methods.

Gut microbiome and metabolome

We will use real-time qPCR to analyse the functional microbial genes responsible for synthesising butyrate[35] and secondary BA.[36] The effect of our intervention on faecal and mucosal-attached microbes associated with inflammation and neoplastic transformation will also be assessed.[37–41]

Global microbiota sequencing will be performed using 16S rRNA gene (16S) sequencing. Genomic DNA extractions will be performed using a bead-beating approach (Qiagen DNeasy Powersoil Kit). Reagent blanks will be included as negative controls, and both cells and genomic DNA from a microbial community of known composition (ZymoBiomics Microbial Community Standards; Zymo Research, Irvine, California, USA) will be included as positive controls. The V4 region of the 16S rRNA gene will be amplified with inline barcoded primers[42] and sequenced on an Illumina MiSeq platform. Sequences will be deconvolved and processed through an in-house sequence quality control pipeline.[43] Taxonomic classification will be performed with the Ribosomal Database Project Naive Bayesian Classifier with the Silva reference database[44 45] for subsequent statistical analyses.[43]

Faecal samples will be transferred to Imperial College London for metabolomics analysis.[1] H Nuclear Magnetic Resonance spectroscopy-based global profiling,[46] together with liquid chromatography-mass spectrometry-based targeted assays (eg, SCFAs, BA, amino acids) will be applied according to in-house developed protocols.[47–50]

Fibre consumption biomarkers

In our studies in Africa and the USA, we have noted an association between fibre intake and the ratio of faecal SCFA to BA. A ratio >10 is associated with low cancer risk, while <5 is associated with high risk ($p<0.0001$). We showed that increasing fibre intake in African Americans was associated with an increase in the ratio from < 5 to >10.[12] The SCFA:BA ratio will be measured monthly in stool samples as a marker of compliance and individual response as described elsewhere.[51 52]

Gastric motility/intestinal transit time

A wireless capsule motility system (SmartPill Corporation, Buffalo, New York, USA) that consists of an indigestible single-use capsule, a receiver, and display software, will be used to examine changes in intestinal transit time. After an overnight fast, the participant will report to the CTSA and swallow the SmartPill capsule with 50 mL of water. The participant is asked to avoid strenuous exercise, alcohol, smoking and medications that may affect GI motility and record bowel movements, food intake, sleep and GI symptoms. The data receiver and diary will be returned after 5 days for analysis. Gastric emptying time, small bowel and colonic transit time are estimated by measuring changes in pH, pressure and temperature. Diet-related changes in motility will provide physiological information about diet response with implications for weight loss success, microbiome changes and mucosal carcinogen contact time.

Mucosal biopsy biomarkers

To assess changes in mucosal biomarkers associated with CRC risk, participants will undergo an unprepped flexible sigmoidoscopy following an overnight fast at various time points indicated in table 1. It is essential to avoid the use of bowel preps as they could affect the microbiota and mucosa. Biopsies will be taken from the sigmoid colon at the furthest region easily accessed (eg, splenic flexure or transverse colon). Participants will be informed that if their sigmoidoscopy procedures reveal any serious health issues, the information gained will be used for research purposes, and participation in the diet intervention will be stopped. A general health assessment by H&E staining and immunohistochemistry to measure epithelial proliferation by Ki67 staining of proliferative cells,[53] epithelial apoptosis by cleaved caspase-3 staining,[54] and inflammation by counting CD3+ intraepithelial lymphocytes and CD68+lamina propria macrophages will be performed as previously reported.[12] Mechanisms of action of the HLD in inducing changes in proliferation will be investigated by measuring the changes in the microbiome and its metabolome and by measuring the changes in the expression of genes known to regulate host defence, inflammation, cell cycling, apoptosis and DNA repair. As funding allows, we will perform supportive investigations into host genome responses to diet by Affymetrix Human Transcriptome Array for expression profiling at the University of Pittsburgh Genomics and Proteomics Laboratories, backed up by in-house RT² qPCR assays (SA biosciences, Qiagen). Also, of interest is the effect of the diets and their relative butyrogenesis on oncogenic miRNAs, which have been associated with increased proliferation in high-meat low-fibre diets, which was reversed by resistant starch fibre supplementation (30 g/day) in human studies.[9 55 56]

Diet assessment/compliance and behaviour questionnaires

Two telephone 24-hour recall interviews will be conducted, 1 weekday and 1 weekend at each time point in table 1. The interviews will be conducted using a multiple-pass interview with the automated self-administered 24-hour diet recall method.[57] We will use questionnaires designed to assess aspects of eating behaviour that may be important for weight management, including diet satisfaction, reasons for meal termination, and hunger and satiety reported about meals and the overall day. The

28-item Diet Satisfaction Questionnaire[58] and the 31-item Reasons Individuals Stop Eating Questionnaire[59] will be used at BL, 6 and 12 months. Visual Analogue Scales[60] allow participants to mark their responses to questions related to hunger and satiety on a line anchored at each end. We will use this tool to evaluate hunger and satiety before and after each meal (six times/day) and at the end of the day.

Physical activity

We will monitor changes in activity level at time points indicated in table 1 to evaluate if group differences could influence weight loss results. Activity will be assessed across several domains (eg, leisure, domestic and exercise) using the leisure time activity survey from the Cancer Prevention Study-3, which provides a continuous indicator of overall activity.[61] Additionally, subjects will be provided with pedometers for use throughout the study and will track and report their steps for 7 days at the time points indicated in table 1. We will ask that participants not make significant changes to physical activity; however, given the study's duration, physical activity changes may occur.

Adherence

Along with regular meal provision, adherence will be encouraged and monitored through regular contact with the study nutritionists, food record sheets, regular weigh-ins to promote self-monitoring, diet recalls and assessment of fibre intake through calculation of the SCFA:BA ratio.

Data management

The study master database and backup procedures will be designed during the initial research phase. Data will be entered into Research Electronic Data Capture (REDCap)[62] and sent to the password-protected Emory data manager to organise, compile and clean for statistical analysis. Data transfer will use secure internet protocols and will observe all IRB and Health Insurance Portability and Accountability Act (HIPAA) requirements.

Withdrawal of participants

At any time during the run-in or throughout the study, participants may withdraw by providing the principal investigator with a written and dated notice of that decision. If they leave the study before the final planned study visit, the researchers may ask the participants to report their weights at the time of their originally scheduled visit and to provide the results related to a colonoscopy they might undergo during the duration of the study. Should participants withdraw from the research without their consent (due to pregnancy, significant health issues, etc), they would be notified by the principal investigator or study coordinator. We will not institute formal withdrawal based on recurrent adenoma, as these are not life-threatening events.

Potential risk and benefits to participants

Given their increased risk of CRC, those diagnosed with adenomatous polyps are more likely to benefit from health promotion programmes.[63] Foodborne illness is a potential risk for participants; however, our strict inclusion and exclusion criteria and food safety protocol will help minimise the likelihood of occurrence. Also, the high-fibre diet may not be well tolerated in some participants (bloating, flatulence), however, we will increase fibre intake gradually and encourage participants to consume adequate amounts of water. Additional risks include venipuncture and gastrointestinal bleeding after mucosal biopsy, although this occurs at rates of less than 1%.[64]

Statistical considerations

Power

For weight loss, power is based on a difference in the trajectory of weight loss between groups over 6 and 12 mos. A 5% decrease in weight is clinically significant due to improvements in blood pressure, lipid profile and insulin sensitivity.[65] A significant difference in weight loss between the two groups is considered to be 1.0 kg±0.9 kg. We expect to observe a larger difference at six mos. (eg, 1.5 kg) and a 1.0 kg difference maintained at 12 months. With a final sample of 60 participants, we would have >95% power for a weight loss difference of 1.0 kg or even 0.8 kg at both time points. If weight change in men differs from women by 0.8 or 0.9 kg (±1.27 kg), the power to detect the difference is 67% or 77%, respectively. These numbers are more than sufficient to observe clinically meaningful changes in secondary outcomes. They will also facilitate our global microbiome and metabolic analyses where exact power calculations are not feasible. Based on our previous research, we plan to enrol and randomise 70 participants, allowing for a drop-out rate of ~15%. Additionally, we estimated the power for detecting a difference in mucosal proliferation at month six between the legume-intervention and control groups based on mucosal proliferation (measured by Ki-67 protein), using changes observed in a 2-week fibre supplementation study with African Americans.[12] While a 20% difference between fibre and placebo implies clinical significance, we have at least 80% power to detect a reduction of at least 9.2%. Our anticipated reduction is higher, approximately 10%.

Proposed analyses

Primary analyses will be intention to treat. Differences at baseline between treatment groups for key variables will be assessed, and those that differ meaningfully will be included as suspected confounders in multivariable models. For those who have a missing outcome of interest, we will fit regression models using those without the missing outcome, use multiple imputations and retain them in the analyses.[66 67] Analyses will be performed using the most recent versions of SAS version 9.4 (SAS Institute, Cary, NC) for body size, blood measures, and transit time

and R (R Foundation) for the mucosal biomarkers and microbiome. Statistical significance will require p<0.05.

To evaluate the role of the HLD on weight reduction, we will focus on the longitudinal change in measures of body size during the most intensive intervention phases (BL to 6 months). Primary analyses will initially use the net change in weight and contrast the change at 6 months in the intervention with that in the control group. We will also evaluate longitudinal changes in mucosal biomarkers from baseline to month 6. Secondarily, we will evaluate weight maintenance and tissue marker changes by contrasting intervention-control changes in measures between 6 and 12 months. Subsequently, we will model outcome trajectories throughout the study using mixed linear models with fixed group, time, group-by-time terms, and random subject and time (slope) terms.

For mucosal biomarkers, after appropriate transformation, changes in markers will be assessed by generalised linear models. Since the genes of interest have been specified in advance, control of the false discovery rate in the stated analyses is not necessary. Exploratory analyses will use the method of Benjamini and Hochberg.[68]

We will assess intervention-related inflammation and insulin sensitivity changes by comparing the area under the curve.[69] In secondary analyses, these results will be analysed using mixed linear models. Mechanisms that may influence the response to the high-legume intervention, such as a change in intestinal transit time and microbiota profiles will also be investigated.

To assess the effect of changes in microbiota on 16S RNAs gene-based analysis of gut microbial composition, the taxonomic profiles will be evaluated with three quantitative approaches that each account for the compositional nature of the data[70]: distance based, abundance based and distribution based. Each approach will be used to understand the microbiota descriptively and to hypothesis–test associations with the clinical data. Intersample distances (eg, beta diversity) will be used to identify biome types ('enterotypes') through hierarchical clustering and the elucidation of taxa influencing sample differentiation. Models will be fit using the 16S profiles with distance-based, (PERMANOVA)[71] to associate the microbiota as a response, with clinical variables as predictors. Multinomial log-linear modelling will also be used to associate clinical variables with hierarchical biome clusters to identify clinical phenotypes. An abundance-based approach applies the additive log ratio (ALR) transformation to relative abundances (to break the spurious correlation among taxa in compositional data) to analyse each taxon as independent and normally distributed.[72] ALR transformed taxonomic abundances will be used as either predictors (multiple regression) or responses (multivariate regression) while controlling for appropriate covariates (eg, sex, age, BMI) in linear models. Based on their positive or negative correlation, ALR taxonomic values will be analysed for correlation, to identify taxa with potential for cooperation (eg, facilitation or syntrophy) or competition (eg, displacement or predation). Distribution-based approaches, such as within-sample or alpha-diversity, analyse the sample profiles as probability distributions. As an index, taxonomic diversity will be analysed in separate linear models as either a response or a predictor variable in association with the clinical variables. In analyses where ALRs or a diversity index is used to represent the microbiome, the microbiome 'as a predictor' will be compared against the inverted model where the microbiome is considered 'as a responder' while controlling for the same covariates. Due to the potentially large number of predictors that may be correlated, principal component (PC) analysis will be applied hierarchically to groups of related variables to identify PCs that can represent the variance for the dataset or identify variables highly correlated with the identified PCs to serve as proxies for the entire dataset. For repeated measures (time), paired difference analyses will be performed to associate predictors with the degree of differences between the two samples from the same subject. Here the difference in taxonomic abundance, overall composition or diversity between the two samples from the same subject is analysed.[73]

Finally, in secondary analyses to estimate the effect of treatment if everyone complied (no non-compliers or dropouts), the group assignment will be used as an instrument using methods described elsewhere.[74] Regression diagnostics, including residual analyses and assessment of autocorrelation patterns, will be used.

Permissions and approvals

The study is also registered with clinicaltrials.gov (NCT04780477). In addition, a material transfer agreement was fully executed.

Study oversight

The Winship Cancer Institute Data and Safety Monitoring Committee (DSMC) is responsible for reviewing pertinent aspects of study conduct, including patient safety, protocol compliance and data collection. Due to the low risk associated with this trial, monitoring will be conducted once within the first year of enrolment for consent and eligibility only. During the initial monitoring visit, 10% of total patient enrolment will be monitored. The DSMC will review monitoring report deficiencies and toxicity data provided by the study team and make recommendations for trial continuation, modification or suspension. The committee reserves the right to conduct additional audits if necessary. The principal investigator or designee is responsible for notifying the DSMC once the trial is open to accrual.

RESULTS DISSEMINATION

We will provide early reports and presentations of the findings to the NIH in progress reports. Because this study will demonstrate important results relevant to populations who consume a westernised diet, we will inform the

scientific community through professional presentations and peer-reviewed journal articles.

Author affiliations
[1]Department of Epidemiology, Rollins School of Public Health and Winship Cancer Institute, Emory University, Atlanta, Georgia, USA
[2]Nutrition and Health Sciences, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA
[3]Department of Medicine, Division of Digestive Diseases, Emory University School of Medicine, Atlanta, Georgia, USA
[4]Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[5]Department of Medicine, Division of Endocrinology, Metabolism and Lipids, Emory University School of Medicine, Atlanta, Georgia, USA
[6]Pulmonary, Allergy and Critical Care Medicine, Center for the Microbiome and Medicine, University of Pittsburg, Pittsburgh, Pennsylvania, USA
[7]Department of Biostatistics and Bioinformatics, Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA
[8]Department of Nutritional Sciences, The Pennsylvania State University, State College, Pennsylvania, USA
[9]Section of Nutrition Research, Division of Digestive Diseases, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Commonwealth Building, Hammersmith Hospital Campus, Imperial College London, South Kensington, London, UK

Acknowledgements The authors wish to express their appreciation to the participants in this study. This study would not be possible without the efforts of the bionutrition staff, nurses and others at the Emory University CTSA. We would also like to thank Amy Ciccarella, RD, LDN, Pennsylvania State University and Halley Fowler, Clinical Research Coordinator, Emory University School of Medicine, Division of Digestive Diseases.

Contributors Content for this manuscript was adapted from the grant proposal and the human subjects research protocol for this project written primarily by TJH and SJDO'K. HH organised the original draft with contributions from all the coauthors. WDF designed the randomisation and contributed to statistical analysis section. JC, AW, TRZ, BLE, BM, BJR, JVL, HH, BC and MRP contributed edits to study operations and specimen acquisition, processing, storage and shipping. All authors were involved in the study design and implementation and reviewed and contributed edits to the drafts of the paper.

Funding Research reported in this publication was supported by the NIH/NCI under award R011CA245063-01A1, by the Intervention Development, Dissemination and Implementation Developing Shared Resource of Winship Cancer Institute under P30CA138292, and by The Georgia Clinical and Translational Science Alliance NIH/NCATS UL1 TR002378 award.

Disclaimer The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Terryl J Hartman http://orcid.org/0000-0001-5942-3161
Annette Wilson http://orcid.org/0000-0002-7426-1786
Barbara J Rolls http://orcid.org/0000-0003-2374-1517
Blaine Loye Eberhart http://orcid.org/0000-0002-6239-053
Jia V Li http://orcid.org/0000-0002-5763-6670
Helaina Huneault http://orcid.org/0000-0003-3866-2387
Ben Cousineau http://orcid.org/0000-0001-8732-2835
Miriam R Perez http://orcid.org/0000-0002-2432-4603

REFERENCES

1 SocietyAC. Key statistics for colorectal cancer 2022, Available: https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html#references
2 Lauby-Secretan B, Scoccianti C, Loomis D, *et al*. Body fatness and cancer--viewpoint of the IARC working group. *N Engl J Med* 2016;375:794–8.
3 O'Keefe SJ. The association between fibre and Hihg-income lifestyle-associated diseases: burkitt's hypothesis revisited. *Lancet Gastroenterol Hepatol* 2019;4:984–96.
4 Cummings JH, Engineer A. Denis burkitt and the origins of the dietary fibre hypothesis. *Nutr Res Rev* 2018;31:1–15.
5 Bingham SA. Mechanisms and experimental and epidemiological evidence relating dietary fibre (non-starch Polysaccharides) and starch to protection against large bowel cancer. *Proc Nutr Soc* 1990;49:153–71.
6 O'Keefe SJD. Diet, microorganisms and their metabolites, and colon cancer. *Nat Rev Gastroenterol Hepatol* 2016;13:691–706.
7 Bobe G, Sansbury LB, Albert PS, *et al*. Dietary flavonoids and colorectal adenoma recurrence in the polyp prevention trial. *Cancer Epidemiol Biomarkers Prev* 2008;17:1344–53.
8 Sreerama YN, Takahashi Y, Yamaki K. Phenolic antioxidants in some vigna species of legumes and their distinct inhibitory effects on alpha glucosidase and pancreatic lipase activities. *J Food Sci* 2012;77:C927–33.
9 Humphreys KJ, Conlon MA, Young GP, *et al*. Dietary manipulation of oncogenic microRNA expression in human rectal mucosa: a randomized trial. *Cancer Prev Res (Phila)* 2014;7:786–95.
10 Bernstein C, Holubec H, Bhattacharyya AK, *et al*. Carcinogenicity of deoxycholate, a secondary bile acid. *Arch Toxicol* 2011;85:863–71.
11 Ocvirk S, O'Keefe SJ. Influence of bile acids on colorectal cancer risk: potential mechanisms mediated by diet - gut microbiota interactions. *Curr Nutr Rep* 2017;6:315–22.
12 O'Keefe SJD, Li JV, Lahti L, *et al*. Fat, fibre and cancer risk in african americans and rural africans. *Nat Commun* 2015;6:6342.
13 Dayib M, Larson J, Slavin J. Dietary fibers reduce obesity-related disorders: mechanisms of action. *Curr Opin Clin Nutr Metab Care* 2020;23:445–50.
14 Hartman TJ, Zhang Z, Albert PS, *et al*. Reduced energy intake and weight loss on a legume-enriched diet lead to improvement in biomarkers related to chronic disease. *Top Clin Nutr* 2011;26:208–15.
15 Hafiz MS, Campbell MD, O'Mahoney LL, *et al*. Pulse consumption improves indices of Glycemic control in adults with and without type 2 diabetes: a systematic review and meta-analysis of acute and long-term randomized controlled trials. *Eur J Nutr* 2022;61:809–24.
16 Rolls BJ, Roe LS, Meengs JS. Reductions in portion size and energy density of foods are additive and lead to sustained decreases in energy intake. *Am J Clin Nutr* 2006;83:11–7.
17 Marciani L, Gowland PA, Spiller RC, *et al*. Gastric response to increased meal viscosity assessed by echo-planar magnetic resonance imaging in humans. *J Nutr* 2000;130:122–7.
18 Gilhooly CH, Das SK, Golden JK, *et al*. Use of cereal fiber to facilitate adherence to a human caloric restriction program. *Aging Clin Exp Res* 2008;20:513–20.
19 McCrory MA, Hamaker BR, Lovejoy JC, *et al*. Pulse consumption, satiety, and weight management. *Adv Nutr* 2010;1:17–30.
20 Chambers ES, Morrison DJ, Frost G. Control of appetite and energy intake by SCFA: what are the potential underlying mechanisms *Proc Nutr Soc* 2015;74:328–36.
21 Chaudhri O, Small C, Bloom S. Gastrointestinal hormones regulating appetite. *Philos Trans R Soc Lond B Biol Sci* 2006;361:1187–209.
22 Marinangeli CPF, Jones PJH. Pulse grain consumption and obesity: effects on energy expenditure, substrate oxidation, body composition, fat deposition and satiety. *Br J Nutr* 2012;108 Suppl 1:S46–51.
23 Saad MJA, Santos A, Prada PO. Linking gut microbiota and inflammation to obesity and insulin resistance. *Physiology (Bethesda)* 2016;31:283–93.

24 Frost G, Sleeth ML, Sahuri-Arisoylu M, *et al*. The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. *Nat Commun* 2014;5:3611.
25 Esteve E, Ricart W, Fernández-Real J-M. Gut microbiota interactions with obesity, insulin resistance and type 2 diabetes: did gut microbiote co-evolve with insulin resistance. *Curr Opin Clin Nutr Metab Care* 2011;14:483–90.
26 Hale VL, Chen J, Johnson S, *et al*. Shifts in the fecal microbiota associated with adenomatous polyps. *Cancer Epidemiol Biomarkers Prev* 2017;26:85–94.
27 Ley RE, Bäckhed F, Turnbaugh P, *et al*. Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* 2005;102:11070–5.
28 Ley RE, Turnbaugh PJ, Klein S, *et al*. Microbial ecology: human gut microbes associated with obesity. *Nature* 2006;444:1022–3.
29 Mifflin MD, St Jeor ST, Hill LA, *et al*. A new predictive equation for resting energy expenditure in healthy individuals. *Am J Clin Nutr* 1990;51:241–7.
30 O'Keefe SJD, Chung D, Mahmoud N, *et al*. Why do african americans get more colon cancer than native africans. *J Nutr* 2007;137(1 Suppl):175S–182S.
31 Lanza E, Hartman TJ, Albert PS, *et al*. High dry bean intake and reduced risk of advanced colorectal adenoma recurrence among participants in the polyp prevention trial. *J Nutr* 2006;136:1896–903.
32 Wheeler ML, Daly A, Evert A, *et al*. Choose your foods: exchange lists for diabetes, sixth edition, 2008: description and guidelines for use. *Journal of the American Dietetic Association* 2008;108:883–8.
33 Hall KD, Kahan S. Maintenance of lost weight and long-term management of obesity. *Med Clin North Am* 2018;102:183–97.
34 Centers for disease control and prevention. In: *National Health and Nutrition Examination Survey (NHANES) Anthropometry Procedures Manual*. Atlanta, GA, 2016.
35 Louis P, Duncan SH, McCrae SI, *et al*. Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J Bacteriol* 2004;186:2099–106.
36 Wells JE, Williams KB, Whitehead TR, *et al*. Development and application of a polymerase chain reaction assay for the detection and enumeration of bile acid 7A-dehydroxylating bacteria in human Feces. *Clin Chim Acta* 2003;331:127–34.
37 Grivennikov SI, Wang K, Mucida D, *et al*. Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth. *Nature* 2012;491:254–8.
38 Kostic AD, Chun E, Robertson L, *et al*. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* 2013;14:207–15.
39 Mima K, Sukawa Y, Nishihara R, *et al*. Fusobacterium nucleatum and T cells in colorectal carcinoma. *JAMA Oncol* 2015;1:653–61.
40 Devkota S, Wang Y, Musch MW, *et al*. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. *Nature* 2012;487:104–8.
41 Attene-Ramos MS, Wagner ED, Plewa MJ, *et al*. Evidence that hydrogen sulfide is a genotoxic agent. *Mol Cancer Res* 2006;4:9–14.
42 Kozich JJ, Westcott SL, Baxter NT, *et al*. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl Environ Microbiol* 2013;79:5112–20.
43 Li K, Epperly MW, Barreto GA, *et al*. Longitudinal fecal microbiome study of total body irradiated mice treated with radiation mitigators identifies bacterial associations with survival. *Front Cell Infect Microbiol* 2021;11:715396.
44 Cole JR, Wang Q, Cardenas E, *et al*. The ribosomal database project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 2009;37:D141–5.
45 Quast C, Pruesse E, Yilmaz P, *et al*. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;41:D590–6.
46 Dona AC, Jiménez B, Schäfer H, *et al*. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Anal Chem* 2014;86:9887–94.
47 Lozupone C, Hamady M, Knight R. Unifrac--an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* 2006;7:371.
48 Lozupone CA, Hamady M, Kelley ST, *et al*. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* 2007;73:1576–85.
49 McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013;8:e61217.
50 Sarafian MH, Lewis MR, Pechlivanis A, *et al*. Bile acid profiling and quantification in biofluids using ultra-performance liquid chromatography tandem mass spectrometry. *Anal Chem* 2015;87:9662–70.
51 Ou J, Carbonero F, Zoetendal EG, *et al*. Diet, microbiota, and microbial metabolites in colon cancer risk in rural africans and african americans. *Am J Clin Nutr* 2013;98:111–20.
52 Park H, Kim M, Kwon GT, *et al*. A high-fat diet increases angiogenesis, solid tumor growth, and lung metastasis of Ct26 colon cancer cells in obesity-resistant BALB/C mice. *Mol Carcinog* 2012;51:869–80.
53 Wood CE, Hukkanen RR, Sura R, *et al*. Scientific and regulatory policy committee (SRPC) review: interpretation and use of cell proliferation data in cancer risk assessment. *Toxicol Pathol* 2015;43:760–75.
54 Gown AM, Willingham MC. Improved detection of apoptotic cells in archival paraffin sections: immunohistochemistry using antibodies to cleaved caspase 3. *J Histochem Cytochem* 2002;50:449–54.
55 Cummins JM, He Y, Leary RJ, *et al*. The colorectal microRNAome. *Proc Natl Acad Sci U S A* 2006;103:3687–92.
56 Humphreys KJ, Cobiac L, Le Leu RK, *et al*. Histone deacetylase inhibition in colorectal cancer cells reveals competing roles for members of the oncogenic miR-17-92 cluster. *Mol Carcinog* 2013;52:459–74.
57 Subar AF, Kirkpatrick SI, Mittl B, *et al*. The automated self-administered 24-hour dietary recall (Asa24): a resource for researchers, clinicians, and educators from the national cancer institute. *J Acad Nutr Diet* 2012;112:1134–7.
58 James BL, Loken E, Roe LS, *et al*. Validation of the diet satisfaction questionnaire: a new measure of satisfaction with diets for weight management. *Obes Sci Pract* 2018;4:506–14.
59 Cunningham PM, Roe LS, Hayes JE, *et al*. Development and validation of the reasons individuals stop eating questionnaire (RISE-Q): a novel tool to characterize Satiation. *Appetite* 2021;161.
60 Flint A, Raben A, Blundell JE, *et al*. Reproducibility, power and validity of visual analogue scales in assessment of appetite sensations in single test meal studies. *Int J Obes Relat Metab Disord* 2000;24:38–48.
61 Rees-Punia E, Matthews CE, Evans EM, *et al*. Reliability and validity of the cancer prevention Study-3 physical activity survey items. *Journal for the Measurement of Physical Behaviour* 2019;2:157–65.
62 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.
63 Turner-McGrievy G, Mandes T, Crimarco A. A plant-based diet for overweight and obesity prevention and treatment. *J Geriatr Cardiol* 2017;14:369–74.
64 Shiffman ML, Farrel MT, Yee YS. Risk of bleeding after endoscopic biopsy or polypectomy in patients taking aspirin or other NSAIDS. *Gastrointest Endosc* 1994;40:458–62.
65 Blackburn G. Effect of degree of weight loss on health benefits. *Obes Res* 1995;3 Suppl 2:211s–6s.
66 Pedersen AB, Mikkelsen EM, Cronin-Fenton D, *et al*. Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol* 2017;9:157–66.
67 Rubin DB. Multiple imputation after 18+ years. *Journal of the American Statistical Association* 1996;91:473–89.
68 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 1995;57:289–300.
69 Matthews JN, Altman DG, Campbell MJ, *et al*. Analysis of serial measurements in medical research. *BMJ* 1990;300:230–5.
70 Gloor GB, Macklaim JM, Pawlowsky-Glahn V, *et al*. Microbiome datasets are compositional: and this is not optional. *Front Microbiol* 2017;8:2224.
71 Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 2001;26:32–46.
72 Tarabichi Y, Li K, Hu S, *et al*. The administration of intranasal live attenuated influenza vaccine induces changes in the nasal microbiota and nasal epithelium gene expression profiles. *Microbiome* 2015;3:74.
73 Stapleton AL, Shaffer AD, Morris A, *et al*. The Microbiome of pediatric patients with chronic rhinosinusitis. *Int Forum Allergy Rhinol* 2021;11:31–9.
74 Nagelkerke N, Fidler V, Bernsen R, *et al*. Estimating treatment effects in randomized clinical trials in the presence of non-compliance. *Stat Med* 2000;19:1849–64.