BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-081379
Article Type:	Protocol
Date Submitted by the Author:	25-Oct-2023
Complete List of Authors:	Hartman, Terryl; Imperial College London Christie, Jennifer; Imperial College London Wilson, Annette; University of Pittsburgh, Ziegler, Thomas R.; Imperial College London Methe, Barbara; University of Pittsburgh, Department of Medicine Flanders, W. Dana; Emory University, Rolls, Barbara ; Pennsylvania State University, Department of Nutritional Sciences Loye Eberhart, Blaine; University of Pittsburgh, Medicine Li, Jia ; Imperial College London, Division of Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine Huneault, Helaina; Emory University, Nutrition and Health Sciences Cousineau, Ben; Emory University Perez, Miriam R.; University of Pittsburgh, Department of Medicine O'Keefe, Stephen; University of Pittsburgh School of Medicine
Keywords:	Obesity, NUTRITION & DIETETICS, GASTROENTEROLOGY, ONCOLOGY

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

Corresponding author: Terry Hartman RD, PhD; tjhartm@emory.edu

Authors:
Terryl J. Hartman, PhD, MPH, RD, Department of Epidemiology, Rollins School of Public Health and Winship Cancer Institute, Emory University[1]

Jennifer Christie, MD, Department of Medicine, Division of Digestive Diseases, Emory University School of Medicine[2]

Annette Wilson, PhD, Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, University of Pittsburgh[3]

Thomas R. Ziegler, MD, Department of Medicine, Division of Endocrinology, Metabolism and Lipids, Emory University School of Medicine[4]

Barbara Methé, PhD, Pulmonary, Allergy and Critical Care Medicine, Center for the Microbiome and Medicine, University of Pittsburg[5]

W. Dana Flanders, MD, DSc, MPH, MA, Department of Biostatistics and Bioinformatics, Department of Epidemiology, Rollins School of Public Health, Emory University[6]

Barbara J. Rolls, PhD, Department of Nutritional Sciences, The Pennsylvania State University[7]

Blaine Loye Eberhart II, Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, University of Pittsburgh[8]

Jia V Li, PhD, 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[9]

Helaina Huneault, RD, Nutrition and Health Sciences, Rollins School of Public Health, Emory University[10]

Ben Cousineau, MS, Nutrition and Health Sciences, Rollins School of Public Health, Emory University[11]

Miriam R Pérez, MS, Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, University of Pittsburgh[12]

Stephen J D O'Keefe, MD, Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, University of Pittsburgh[13]

Keywords: Colorectal cancer prevention, adenomatous polyps, obesity, legumes, high fiber diet, gut microbiome, metabolome, gut transit time

Abstract:

Introduction: Recently published studies support the beneficial effects of consuming fiber-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 Fiber-rich Foods to Treat Obesity and Prevent Colon Cancer randomized clinical trial (RCT) will test whether a high fiber diet featuring legumes will simultaneously facilitate weight reduction and suppress colonic mucosal biomarkers of colorectal cancer (CRC).

Methods/design: This study is designed to characterize changes in (1) body weight; (2) biomarkers of insulin resistance and systemic inflammation; (3) compositional and functional profiles of the fecal microbiome and metabolome; (4) mucosal biomarkers of colorectal cancer risk; and (5) gut transit. Approximately 60 overweight or obese adults with a history of noncancerous adenomatous polyps within the previous three years will be recruited and randomized to one of two weight-loss diets. Following a 1-week run-in, participants in the intervention arm will receive pre-portioned high fiber 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 (e.g., 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.

Trial registration: This protocol is registered with the U.S. National Institutes of Health trial registry, ClinicalTrials.gov, under the identifier NCT04780477. First posted March 2nd, 2021; last verified May 16th, 2022.

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 sub-award to the University of Pittsburgh. The study protocol was approved by the Emory Institutional Review Board (IRB approval number: 00000563).

Strengths and limitations of this study

- This will be the first study to measure the effects of a high fiber diet on the human microbiota, metabolome, and colonic mucosal biomarkers of CRC over 12 months of intervention and to assess the effects of nutrition education on obesity and CRC risk at ~3 years
- A novel fecal biomarker of dietary composition, namely SCFA, indicative of fiber intake, together with bile acids, as markers of total fat consumption, will be measured to assess compliance with the dietary intervention and the need for more intense behavior modification and fiber supplementation.
  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.
- Given that this cancer prevention study specifically targets healthy individuals, future results may not be generalizable to cancer patients.

Background

Colorectal cancer (CRC) is the third most common form of cancer in the United States[1]. Obesity increases the risk of at least 13 cancers, including CRC[2, 3]. Burkitt's original hypothesis[4] from 1963 highlights that westernized diseases such as CRC and obesity may result from fiber deficiency from the commercial refinement of foods. Many plausible mechanisms explain why high-fiber diets, especially a high legume diet (HLD), may reduce CRC risk. First, fiber 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]. Secondly, microbiota break down plant cell walls releasing phytochemicals, which also have powerful anti-inflammatory and anti-carcinogenic effects[7, 8]. Thirdly, colonic transit is accelerated, reducing contact time with luminal carcinogens, such as heterocyclic amines formed from cooked red meat,[9] and secondary bile acids, induced by a high-fat diet and synthesized by the colonic microbiota[10, 11].

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

Our prior research suggests that fiber 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, like legumes, should potentiate weight control[16]. Legumes are high in resistant starch, insoluble fiber, and especially soluble fiber. Therefore, legumes absorb water during digestion, increasing viscosity, encouraging stomach distension, and inducing satiation[17].

Fiber-rich diets may also affect other physiological mechanisms important for weight control[18, 19]. Trypsin inhibitors and other bioactive compounds found in legumes (e.g., lectins) may directly stimulate cholecystokinin (CCK) 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 characterize 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 randomized controlled trial of a high legume diet compared to 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 hypothesize that restoring the diet with natural high-fiber content, principally with legumes, will lead to a more significant weight loss and improvements in biomarkers associated with colon cancer risk compared to a control diet. See Table 1 for a detailed timeline of outcome measures.

Methods and Analysis

This study is a parallel arm randomized 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 randomized to. Pre-portioned 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 yrs. old, (2) BMI 25-40 kg/m^2, (3) colonoscopy within three years that found ≥1 adenoma >0.5 cm (4) English speaking, (5) ambulatory, (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 behavior, (6) regular use of medication that may interfere with study procedures, (7) women currently pregnant, breastfeeding, 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 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 has been attained. After the study is completed, all data and specimens will be kept secure according to NIH and FDA regulations.

Study intervention

Run-in phase

Before randomization, participants will proceed through a 1-week run-in where foods representative of the control diet will be provided. The run-in helps to standardize 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 version 4.1.3. to allocate subjects to each treatment arm. To conceal the randomization 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 ~500-1,000 kcal/day)[29].

Prepared portion control entrées

To all participants, we will provide pre-portioned 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 (e.g., navy, pinto, black, kidney beans, etc.) to limit nutrient and phytochemical variability. The diet will contain approximately 250g of legumes per day (~1 ½ cups cooked). This level will add approximately 30 grams of dietary fiber/day from the legume dishes, ensuring a total intake of ~45-50g/day. Previously, this level reduced colonic mucosal biomarkers of cancer risk within two 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 (PPT)[31]. The control group will also receive pre-portioned meal replacement entrées with legumes replaced by lean chicken/meat. All entrées will be prepared, pre-portioned, 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 (e.g., weight checks), replacement of high energy density foods with lower energy density alternatives, and portion control, among other strategies[33]. Skill-building and behavioral strategies to address the aforementioned behaviors 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 behavior 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 behavior 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 Westernized 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.

Table 1.

Intervention and Assessment Activities (W=weekly; M=monthly; B=bi-monthly, 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
Fecal fiber 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[34], DSQ[35]	X						X						X	
VAS[36]	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 & sides at weight maintenance energy level is provided to all participants before baseline assessments and randomization. Ext = extended follow-up (when a participant has a subsequent colonoscopy)

Data Collection

Anthropometry

Bodyweight, height, and waist circumference (WC) will be measured using the most recent NHANES procedures[37]. 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 (DXA).

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 our Supplementary Materials for further details on these methods.

Stool and urine sample collection

Procedures developed in our NIH-supported studies will be used to ensure scientific rigor for collecting and analyzing 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 our Supplementary Materials for further details on these methods.

Gut microbiome and metabolome

We will use real-time qPCR to analyze the functional microbial genes responsible for synthesizing butyrate[38] and secondary bile acids[39]. The effect of our intervention on fecal and mucosal-attached microbes associated with inflammation and neoplastic transformation will also be assessed[40-44].

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, CA) will be included as positive controls. The V4 region of the 16S rRNA gene will be amplified with inline barcoded primers[45] and sequenced on an Illumina MiSeq platform. Sequences will be deconvolved and processed through an in-house sequence quality control pipeline[46]. Taxonomic classification will be performed with the Ribosomal Database Project Naive Bayesian Classifier with the Silva reference database[47, 48] for subsequent statistical analyses[46].

Fecal samples will be transferred to Imperial College London for metabolomics analysis. ^{1}H Nuclear Magnetic Resonance (NMR) spectroscopy-based global profiling[49], together with Liquid Chromatography-Mass Spectrometry (LC-MS)-based targeted assays (e.g., SCFAs, bile acids, amino acids) will be applied according to in-house developed protocols[50-53].

Fiber Consumption Biomarkers

In our studies in Africa and the US, we have noted an association between fiber intake and the ratio of fecal SCFA to bile acids (BA). A ratio >10 is associated with low cancer risk, while <5 is associated with high risk ($p<0.0001$). We showed that increasing fiber 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[54, 55].

Gastric Motility/Intestinal Transit Time

A wireless capsule motility system (SmartPill Corporation, Buffalo, NY) 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 five 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 physiologic 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 (e.g., 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[56], epithelial apoptosis by cleaved caspase-3 staining[57], 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 defense, 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-fiber diets, which was reversed by resistant starch fiber supplementation (30 g/day) in human studies[9, 58, 59].

Diet Assessment/Compliance and Behavior Questionnaires

Two telephone 24-hour recall interviews will be conducted, one weekday and one weekend at each time point in Table 1. The interviews will be conducted using a multiple-pass interview with the Automated Self-Administered (ASA) 24-hour diet recall method[60]. We will use questionnaires designed to assess aspects of eating behavior 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 (DSQ)[35] and the 31-item Reasons Individuals Stop Eating Questionnaire (RISE-Q)[34] will be used at BL, 6, and 12 months. Visual analog scales (VAS)[36] 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 (6 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 (e.g., leisure, domestic, exercise) using the leisure time activity survey from the Cancer Prevention Study-3

(CPS-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 seven 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 fiber 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 organize, compile, and clean for statistical analysis. Data transfer will use secure internet protocols and will observe all IRB and 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 participant 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 programs[63]. Foodborne illness is a potential risk for participants; however, our strict inclusion and exclusion criteria and food safety protocol will help minimize the likelihood of occurrence. Also, the high-fiber diet may not be well tolerated in some participants (bloating, flatulence), however we will increase fiber 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 clinically significant difference in weight loss is assumed to be 1.0 kg $\pm$ 0.9 kg[65]. We expect to observe a larger difference at 6 mos. (e.g., 1.5 kg) and a 1.0 kg difference maintained at 12 mos. 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 ($\pm$ 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 enroll and randomize 70 participants, allowing for a dropout 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 fiber supplementation study with African Americans[12]. While a 20% difference between fiber 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 intent-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 (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 six 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 generalized linear models, Since the genes of interest have been specified in advance, control of the false discovery rate (FDR) 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 analyzed 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. Inter-sample distances (e.g. 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 modeling 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 analyze 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 (e.g., sex, age, BMI) in linear models. Based on their positive or negative correlation, ALR taxonomic values will be analyzed for correlation, to identify taxa with potential for cooperation (e.g., facilitation or syntrophy) or competition (e.g., displacement or predation). Distribution-based approaches, such as within-sample or alpha-diversity, analyze the sample profiles as probability distributions. As an index, taxonomic diversity will be analyzed 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 analysis (PCA) will be applied hierarchically to groups of related variables to identify principal components (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 analyzed[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 enrollment for consent and eligibility only. During the initial monitoring visit, 10% of total patient enrollment 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 westernized 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
[2]Department of Medicine, Division of Digestive Diseases, Emory University School of Medicine
[3,8,12,13]Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, University of Pittsburgh
[4]Department of Medicine, Division of Endocrinology, Metabolism and Lipids, Emory University School of Medicine
[5]Pulmonary, Allergy and Critical Care Medicine, Center for the Microbiome and Medicine, University of Pittsburg
[6]Department of Biostatistics and Bioinformatics, Department of Epidemiology, Rollins School of Public Health, Emory University
[7]Department of Nutritional Sciences, The Pennsylvania State University
[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
[10,11]Nutrition and Health Sciences, Rollins School of Public Health, Emory University

Acknowledgments 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 TH and SO. HH organized the original draft with contributions from all the co-authors. WDF designed the randomization and contributed to statistical analysis section. JC, AW, TZ, LE, BM, BR, JL, HH, BC and MP 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 Intervention Development, Dissemination, and Implementation Developing Shared Resource of Winship Cancer Institute of Emory University and NIH/NCI under award number P30CA138292. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Georgia Clinical and Translational Science Alliance NIH/NCATS UL1 TR002378 award also contributed to the support of this research.

Competing interests No competing interests

Patient consent Not required

Ethics approval The study protocol was approved by the Emory Institutional Review Board (IRB approval number: 00000563).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Data will be shared in four ways: 1) preliminary and final results will be shared in the traditional conference presentation and manuscript formats; 2) results will be summarized and presented to the lay public via the NIH-supported CTSI listserv, news releases, and community lectures, as well as relevant entities at collaborator sites; 3) results will be summarized and presented via lectures and review articles to the academic community; 4) actual study data in raw and summary form, with protected health information removed, as well as access to the study database, will be available for reference of qualified professional colleagues and auditors up to 10 years after study publication. Data will be deposited in appropriate research data repositories (e.g., NIH Common Fund Metabolomics Workbench).

ORCID ID's Annette Wilson: 0000-0002-7426-1786
Helaina E. Huneault: 0000-0003-3866-2387
Jia V. Li: 0000-0002-5763-6670
Terryl J. Hartman: 0000-0001-5942-3161
Miriam R Pérez: 0000-0002-2432-4603
Ben Cousineau: 0000-0001-8732-2835
Barbara J. Rolls: 0000-0003-2374-1517
Blaine Loye Eberhart: 0000-0002-6239-053

References

1. Society AC. Key statistics for colorectal cancer 2022, January 12 [Available from: 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(8):794-8.
3. O'Keefe SJ. The association between fibre and hihg-income lifestyle-associated diseases: Burkitt's hypothesis revisited. *Lancet Gastro Hep*. 2019;4:in press.
4. Cummings JH, Engineer A. Denis Burkitt and the origins of the dietary fibre hypothesis. *Nutr Res Rev*. 2018;31(1):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(2):153-71.
6. O'Keefe SJ. Diet, microorganisms and their metabolites, and colon cancer. *Nat Rev Gastroenterol Hepatol*. 2016;13(12):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(6):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(9):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(8):786-95.
10. Bernstein C, Holubec H, Bhattacharyya AK, et al. Carcinogenicity of deoxycholate, a secondary bile acid. *Arch Toxicol*. 2011;85(8):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(4):315-22.
12. O'Keefe SJ, 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(6):445-50.
14. Hartman TJ, Zhang Z, Albert P, 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(2):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(1):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(1):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(6):513-20.
19. McCrory MA, Hamaker BR, Lovejoy JC, et al. Pulse consumption, satiety, and weight management. *Adv Nutr*. 2010;1(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(3):328-36.
21. Chaudhri O, Small C, Bloom S. Gastrointestinal hormones regulating appetite. *Philos Trans R Soc Lond B Biol Sci*. 2006;361(1471):1187-209.
22. Marinangeli CP, Jones PJ. 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 MJ, Santos A, Prada PO. Linking Gut Microbiota and Inflammation to Obesity and Insulin Resistance. *Physiology (Bethesda)*. 2016;31(4):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, Fernandez-Real JM. 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(5):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(1):85-94.
27. Ley RE, Backhed F, Turnbaugh P, et al. Obesity alters gut microbial ecology. *Proc Natl Acad Sci*. 2005;102(31):11070-5.
28. Ley RE, Turnbaugh PJ, Klein S, et al. Microbial ecology: Human gut microbes associated with obesity. *Nature*. 2006;444(7122):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(2):241-7.
30. O'Keefe SDC, D Mahmoud, N, Sepulveda AR, and Manafe M. Why Do African Americans Get More Colon Cancer than Native Africans? *J Nutr*. 2007;137(1):175S-82.
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(7):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. *J Am Diet Assoc*. 2008;108(5):883-8.
33. Hall KD, Kahan S. Maintenance of Lost Weight and Long-Term Management of Obesity. *Med Clin North Am*. 2018;102(1):183-97.
34. 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;161:105127.
35. James BL, Loken E, Roe LS, et al. Validation of the Diet Satisfaction Questionnaire: a new measure of satisfaction with diets for weight management. *Obesity science & practice*. 2018;4(6):506-14.
36. 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(1):38-48.
37. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey (NHANES) Anthropometry Procedures Manual. Atlanta, GA; 2016.
38. 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(7):2099-106.
39. 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 7α-dehydroxylating bacteria in human feces. *Clin Chim Acta*. 2003;331(1-2):127-34.
40. 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(7423):254-8.

41. Kostic AD, Chun E, Robertson L, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe*. 2013;14(2):207-15.
42. Mima K, Sukawa Y, Nishihara R, et al. Fusobacterium nucleatum and T Cells in Colorectal Carcinoma. *JAMA Oncol*. 2015;1(5):653-61.
43. Devkota S, Wang Y, Musch MW, et al. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. *Nature*. 2012;487(7405):104-8.
44. Attene-Ramos MS, Wagner ED, Plewa MJ, et al. Evidence That Hydrogen Sulfide Is a Genotoxic Agent. *Mol Cancer Res*. 2006;4(1):9-14.
45. 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(17):5112-20.
46. 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.
47. 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(Database issue):D141-5.
48. 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(Database issue):D590-6.
49. 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(19):9887-94.
50. Lozupone C, Hamady M, Knight R. UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics*. 2006;7:371.
51. 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(5):1576-85.
52. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8(4):e61217.
53. 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(19):9662-70.
54. 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(1):111-20.
55. 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(11):869-80.
56. 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(6):760-75.
57. 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(4):449-54.
58. Cummins JM, He Y, Leary RJ, et al. The colorectal microRNAome. *Proc Natl Acad Sci U S A*. 2006;103(10):3687-92.
59. 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(6):459-74.

60. 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(8):1134-7.
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(3):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(2):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(5):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(4):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(434):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* 1995;57(1):298-310.
69. Matthews JNS, Atlman DG, Campbell JJ, et al. Analysis of serial measurements in medical research. *Br Med J*. 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(1):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(1):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.

Data Collection – Supplementary Materials:

1) **Blood collection:** Approximately 50 ml of blood will be collected from the subjects at baseline, 6 and 12 months, and at extended follow-up. Samples will be collected after an overnight fast and in the early morning. Serum and plasma samples will be aliquoted in smaller vials, frozen at –80°C, and stored for later analysis of insulin, glucose, and CRP using standard techniques. Medication use will be assessed at each collection.

2) **Insulin and Insulin resistance measures:** As mentioned above, fasting blood samples will be collected, aliquoted and stored frozen at -80°C for analysis at study completion. Insulin resistance will be evaluated by the Homeostasis Assessment Model from fasting insulin and glucose [= fasting insulin (μU/mL) × fasting glucose (mmol/L)/22.5]; values > 2.61 will be considered insulin resistant [1]. Samples will be analyzed under the direction of Dr. Ngoc-Anh Le, Biomarker Core Laboratory at the Atlanta VA Medical Center. Glucose will be determined by colorimetric methods (Sekisui Diagnostics, Exton, PA), plasma insulin levels assessed using the immunoturbidometric method (Sekisui), and high sensitivity CRP analyzed via sandwich enzyme immunoassay (ALPCO). Samples will be grouped in random order for analysis, and a 10% blind quality control included. The CV for these analyses in the Le lab ranges between 3.9%-6.1%.

3) **Urine Collection:** 50 ml urine samples will be collected at baseline, 6 and 12 months and at extended follow-up in plastic containers by clean-catch technique, transported, and stored in the same way as fecal samples. Samples will be aliquoted into five containers/time point, coded by GCRC laboratory staff, and held at –80°C for future metabolome analysis led by Dr. O'Keefe [2].

4) **Fecal Collection:** Stool samples will be collected during the study. Subjects will be instructed in the use of a plastic device to cover the toilet seat and collect the stool. Two separate ~5 g samples will be taken for fecal microbiome analyses (mechanistic studies) and two additional samples for fecal SCFA and bile acid analyses (markers of compliance with HLD) at the time points specified in Table 1. All samples will be transported to the laboratory, coded by the research assistant, and held at –80° C for future DNA extraction and microbiome analysis by targeted and global approaches led by the O'Keefe laboratory.

5) Body composition: Body composition will be assessed using dual energy X-ray absorptiometry (DXA)[3] in the GCRC at Emory University. This method uses a whole-body scanner to measure total body composition and fat content with a high degree of precision. It is safe and noninvasive with little burden to the individual. Data from the DXA scans will be used to assess longitudinal changes in body fat that accompany weight loss. This data will also allow examination of changes in fat distribution at defined regions in the body. New software allows for the estimation of visceral fat. Women who could potentially become pregnant will be given a pregnancy test prior to a DXA.

References:

1. Monzillo LU, Hamdy O. Evaluation of insulin sensitivity in clinical practice and in research settings. *Nutr Rev*. 2003;61(12):397-412.
2. O'Keefe SJ, Li JV, Lahti L, et al. Fat, fibre and cancer risk in African Americans and rural Africans. *Nat Commun*. 2015;6:6342.
3. Lunar GH. X-ray bone densitometer with enCORE v17 software—user manual. *Madison: GE Healthcare Lunar*. 2016.

Meal Plan A – Healthy American Control group meal plan

Emory GCRC Bionutrition

Meal	1	2	3	4	5	6	7
Breakfast							
Lunch	Lemon Chicken Potato Casserole	Quinoa Tofu Pesto	Turkey Burger on a roll (side of: lettuce, onion, catsup)	South of the Border Chicken Stew	Bulgur-Beef Meatballs with Spaghetti & Sauce	Lemon Chicken w/ Bulgur	Spinach-Cheese Noodles
Sides	*Non-Starchy Vegetables (i.e. Roasted Brussel Sprouts), Fruit (i.e. Pear)*	*Non-Starchy Vegetable (i.e. Steamed Asparagus), Fruit (i.e. Orange)*	*Non-Starchy Vegetable (i.e. Sliced Tomatoes), Fat (i.e. Avocado)*	*Starch (i.e. Whole Grain Sliced Bread), Fat (i.e. Avocado)*	*Vegetable (i.e. Salad)*	*Non-Starchy Vegetable (i.e. Green Beans), Fruit (i.e. Nectarine)*	*Protein (i.e. Grilled Salmon), Non-Starchy Vegetable + Combination Food [Carbohydrate & Fat] (i.e. Raw Sweet Peppers & Hummus)*
Dinner	Pasta & Peanut Salad	Stovetop BBQ Chicken with basmati rice & raw spinach	Spaghetti with Turkey Meat Sauce	Turkey Meatloaf	Spiced Chicken & Rice	Potato Green Bean Bake	Beef & Broccoli Stir Fry with Basmati Rice
Sides	*Starch (i.e. Pretzels), Fruit (i.e. Tangerine)*	*Starchy Vegetable (i.e. Corn on the Cob)*	*Starchy Vegetable (i.e. Mixed Vegetable, frozen), Fruit (i.e. Pear)*	*Starch Vegetable (i.e. Baked Potato), Vegetable (i.e. Salad)*	*Non-Starchy Vegetable (i.e. Steamed Broccoli), Fat (i.e. Almonds)*	*Vegetable (i.e. Salad), Non-Starchy Vegetable + Fat (i.e. Celery & Peanut Butter)*	*Fruit (i.e. Seedless Grapes)*

Meal Plan B – Legume Intervention group meal plan Emory GCRC Bionutrition

Meal	1	2	3	4	5	6	7
Breakfast							
Lunch	Chickpea Salad Sandwich (inc. whole grain bread, lettuce)	Lentil Tomato Salad	Black Bean Salad	Black-Eyed Pea Curry w/ Basmati Rice	Chana Massala (flavorful chickpea, tomato & spinach stew) w/ Basmati Rice	Falafel (baked chickpea patties) [served with whole wheat pita, lettuce, tomato, onion, tahini]	Southwest Bean Soup
Sides	*Non-Starchy Vegetable (i.e. Tomato Slices), Fruit (i.e. Apple)*	*Starchy Vegetable (i.e. Baked Sweet Potato), Fruit (i.e. Apple)*	*Combination Food (carbohydrate, fat) [i.e. Tortilla Chips], Non-Starchy Vegetables (i.e. Cherry Tomatoes)*	*Fruit (i.e. Orange)*	*Non-Starchy Vegetable + Combination Food (Carbohydrate & Fat) [i.e. Baby Carrots & Hummus]*	*Non-Starchy Vegetable (i.e. Sliced Cucumber)*	*Vegetable (i.e. Salad), Fruit (i.e. Apple)*
Dinner	Pineapple Lentils w/ Basmati Rice	Quinoa Chili	Bean & Chicken Cassoulet (casserole)	Caldo Verde (Potato, Kale & White Bean Soup)	White Bean Onion Stew	African Peanut Stew	Tamale Pie
Sides	*Vegetable (i.e. Salad)*	*Fat (i.e. Avocado), Fruit (i.e. Peach)*	*Vegetable (i.e. Salad), Non-Starchy Vegetable (i.e. Steamed Broccoli)*	*Starch (i.e. Whole Grain Toast), Medium-Fat Protein (i.e. Poached or Hard-Boiled Egg)*	*Vegetable (i.e. Salad), Medium-Fat Protein (i.e. Hard-Boiled Egg)*	*Non-Starchy Vegetable (i.e. Celery Sticks), Fruit (i.e. Seedless Grapes)*	*Non-Starchy Vegetable (i.e. Baby Carrots), Fruit (i.e. Tangerine)*

SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents*

Section/item	Item No	Description
Administrative information		
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym Page 1
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry Page 2
	2b	All items from the World Health Organization Trial Registration Data Set N/A
Protocol version	3	Date and version identifier Page 2
Funding	4	Sources and types of financial, material, and other support Page 13
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors Page 13
	5b	Name and contact information for the trial sponsor Page 13
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities Page 13
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee) Page 12
Introduction		

Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention Page 4,10
	6b	Explanation for choice of comparators Pages 5-10
Objectives	7	Specific objectives or hypotheses Page 4,5
Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory) Page 4-7
Methods: Participants, interventions, and outcomes		
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained Page 5, 6
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists) Page 5
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered Page 5-7
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease) Page 10
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests) Page 10
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial Page 6

Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended Page 4-9
Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure) Page 7
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations Page 10,11
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size Page 5

Methods: Assignment of interventions (for controlled trials)

Allocation:

Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions Page 6
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned Page 6
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions Page 6
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how Page 5,6

	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial Page 5
Methods: Data collection, management, and analysis		
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol Pages 7-10
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols Page 10
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol Page 10
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol Page 10-12
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses) Page 10-12
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation) Page 10-12
Methods: Monitoring		
Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed Page 12

	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial Page 10,11
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct Page 10
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor Page 10,12

Ethics and dissemination

Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval Page 12,13
Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators) Page 12,13
Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32) Page 5
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable N/A
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial Page 5
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site Page 13
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators Page 13
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation N/A

Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions Page 12
	31b	Authorship eligibility guidelines and any intended use of professional writers Page 13
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code N/A
Appendices		
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates Supplementary Materials
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable Page 7

*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group under the Creative Commons "Attribution-NonCommercial-NoDerivs 3.0 Unported" license.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-081379.R1
Article Type:	Protocol
Date Submitted by the Author:	20-Dec-2023
Complete List of Authors:	Hartman, Terryl; Imperial College London Christie, Jennifer; Imperial College London Wilson, Annette; University of Pittsburgh, Ziegler, Thomas R.; Imperial College London Methe, Barbara; University of Pittsburgh, Department of Medicine Flanders, W. Dana; Emory University, Rolls, Barbara ; Pennsylvania State University, Department of Nutritional Sciences Loye Eberhart, Blaine; University of Pittsburgh, Medicine Li, Jia ; Imperial College London, Division of Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine Huneault, Helaina; Emory University, Nutrition and Health Sciences Cousineau, Ben; Emory University Perez, Miriam R.; University of Pittsburgh, Department of Medicine O'Keefe, Stephen; University of Pittsburgh School of Medicine
Primary Subject Heading:	Nutrition and metabolism
Secondary Subject Heading:	Oncology, Gastroenterology and hepatology
Keywords:	Obesity, NUTRITION & DIETETICS, GASTROENTEROLOGY, ONCOLOGY

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

Corresponding author: Terry Hartman RD, PhD; tjhartm@emory.edu

Authors:
Terryl J. Hartman, PhD, MPH, RD, Department of Epidemiology, Rollins School of Public Health and Winship Cancer Institute, Emory University[1]

Jennifer Christie, MD, Department of Medicine, Division of Digestive Diseases, Emory University School of Medicine[2]

Annette Wilson, PhD, Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, University of Pittsburgh[3]

Thomas R. Ziegler, MD, Department of Medicine, Division of Endocrinology, Metabolism and Lipids, Emory University School of Medicine[4]

Barbara Methé, PhD, Pulmonary, Allergy and Critical Care Medicine, Center for the Microbiome and Medicine, University of Pittsburg[5]

W. Dana Flanders, MD, DSc, MPH, MA, Department of Biostatistics and Bioinformatics, Department of Epidemiology, Rollins School of Public Health, Emory University[6]

Barbara J. Rolls, PhD, Department of Nutritional Sciences, The Pennsylvania State University[7]

Blaine Loye Eberhart II, Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, University of Pittsburgh[8]

Jia V Li, PhD, 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[9]

Helaina Huneault, RD, Nutrition and Health Sciences, Rollins School of Public Health, Emory University[10]

Ben Cousineau, MS, Nutrition and Health Sciences, Rollins School of Public Health, Emory University[11]

Miriam R Pérez, MS, Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, University of Pittsburgh[12]

Stephen J D O'Keefe, MD, Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, University of Pittsburgh[13]

Keywords: Colorectal cancer prevention, adenomatous polyps, obesity, legumes, high fiber diet, gut microbiome, metabolome, gut transit time

Abstract:

Introduction: Recently published studies support the beneficial effects of consuming fiber-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 Fiber-rich Foods to Treat Obesity and Prevent Colon Cancer randomized clinical trial (RCT) will test whether a high fiber diet featuring legumes will simultaneously facilitate weight reduction and suppress colonic mucosal biomarkers of colorectal cancer (CRC).

Methods/design: This study is designed to characterize changes in (1) body weight; (2) biomarkers of insulin resistance and systemic inflammation; (3) compositional and functional profiles of the fecal microbiome and metabolome; (4) mucosal biomarkers of colorectal cancer risk; and (5) gut transit. Approximately 60 overweight or obese adults with a history of noncancerous adenomatous polyps within the previous three years will be recruited and randomized to one of two weight-loss diets. Following a 1-week run-in, participants in the intervention arm will receive pre-portioned high fiber 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 (e.g., 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.

Trial registration: This protocol is registered with the U.S. National Institutes of Health trial registry, ClinicalTrials.gov, under the identifier NCT04780477. First posted March 2nd, 2021; last verified May 16th, 2022.

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 sub-award to the University of Pittsburgh. The study protocol was approved by the Emory Institutional Review Board (IRB approval number: 00000563).

Strengths and limitations of this study

- This study offers a comprehensive analysis of the effects of a high fiber diet on human microbiota, metabolome, and colonic mucosal biomarkers of 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 fecal short-chain fatty acids (SCFAs), indicative of fiber consumption, and bile acids, representative of fat intake, aids in monitoring dietary adherence and determining if intensified behavioral changes and fiber 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 generalizable to cancer patients.

Background

Colorectal cancer (CRC) is the third most common form of cancer in the United States[1]. Obesity increases the risk of at least 13 cancers, including CRC[2, 3]. Burkitt's original hypothesis[4] from 1963 highlights that westernized diseases such as CRC and obesity may result from fiber deficiency from the commercial refinement of foods. Many plausible mechanisms explain why high-fiber diets, especially a high legume diet (HLD), may reduce CRC risk. First, fiber 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]. Secondly, microbiota break down plant cell walls releasing phytochemicals, which also have powerful anti-inflammatory and anti-carcinogenic effects[7, 8]. Thirdly, colonic transit is accelerated, reducing contact time with luminal carcinogens, such as heterocyclic amines formed from cooked red meat,[9] and secondary bile acids, induced by a high-fat diet and synthesized by the colonic microbiota[10, 11].

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

Our prior research suggests that fiber 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, like legumes, should potentiate weight control[16]. Legumes are high in resistant starch, insoluble fiber, and especially soluble fiber. Therefore, legumes absorb water during digestion, increasing viscosity, encouraging stomach distension, and inducing satiation[17].

Fiber-rich diets may also affect other physiological mechanisms important for weight control[18, 19]. Trypsin inhibitors and other bioactive compounds found in legumes (e.g., lectins) may directly stimulate cholecystokinin (CCK) 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 characterize 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 randomized controlled trial of a high legume diet compared to 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 hypothesize that restoring the diet with natural high-fiber content, principally with legumes, will lead to a more significant weight loss and improvements in biomarkers associated with colon cancer risk compared to a control diet. See Table 1 for a detailed timeline of outcome measures.

Methods and Analysis

This study is a parallel arm randomized 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 randomized to. Pre-portioned 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 yrs. old, (2) BMI 25-40 kg/m^2, (3) colonoscopy within three years that found/removed ≥1 adenoma >0.5 cm, (4) English speaking, (5) ambulatory, (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 behavior, (6) regular use of medication that may interfere with study procedures, (7) women currently pregnant, breastfeeding, 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 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 has been attained. After the study is completed, all data and specimens will be kept secure according to NIH and FDA regulations.

Study intervention

Run-in phase

Before randomization, participants will proceed through a 1-week run-in where foods representative of the control diet will be provided. The run-in helps to standardize 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 version 4.1.3. to allocate subjects to each treatment arm. To conceal the randomization 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 ~500-1,000 kcal/day)[29].

Prepared portion control entrées

To all participants, we will provide pre-portioned 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 (e.g., navy, pinto, black, kidney beans, etc.) to limit nutrient and phytochemical variability. The diet will contain approximately 250g of legumes per day (~1 ½ cups cooked). This level will add approximately 30 grams of dietary fiber/day from the legume dishes, ensuring a total intake of ~45-50g/day. Previously, this level reduced colonic mucosal biomarkers of cancer risk within two 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 (PPT)[31]. The control group will also receive pre-portioned meal replacement entrées with legumes replaced by lean chicken/meat. Please see our Supplementary Materials for example meal plans for each group. All entrées will be prepared, pre-portioned, 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 (e.g., weight checks), replacement of high energy density foods with lower energy density alternatives, and portion control, among other strategies[33]. Skill-building and behavioral strategies to address the aforementioned behaviors 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 behavior 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 behavior 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 Westernized 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.

Table 1.

Intervention and Assessment Activities (W=weekly; M=monthly; B=bi-monthly, 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
Fecal fiber 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[34], DSQ[35]	X						X						X	
VAS[36]	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 & sides at weight maintenance energy level is provided to all participants before baseline assessments and randomization. Ext = extended follow-up (when a participant has a subsequent colonoscopy)

Data Collection

Anthropometry

Bodyweight, height, and waist circumference (WC) will be measured using the most recent NHANES procedures[37]. 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 (DXA).

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 our Supplementary Materials for further details on these methods.

Stool and urine sample collection

Procedures developed in our NIH-supported studies will be used to ensure scientific rigor for collecting and analyzing 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 our Supplementary Materials for further details on these methods.

Gut microbiome and metabolome

We will use real-time qPCR to analyze the functional microbial genes responsible for synthesizing butyrate[38] and secondary bile acids[39]. The effect of our intervention on fecal and mucosal-attached microbes associated with inflammation and neoplastic transformation will also be assessed[40-44].

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, CA) will be included as positive controls. The V4 region of the 16S rRNA gene will be amplified with inline barcoded primers[45] and sequenced on an Illumina MiSeq platform. Sequences will be deconvolved and processed through an in-house sequence quality control pipeline[46]. Taxonomic classification will be performed with the Ribosomal Database Project Naive Bayesian Classifier with the Silva reference database[47, 48] for subsequent statistical analyses[46].

Fecal samples will be transferred to Imperial College London for metabolomics analysis. ^{1}H Nuclear Magnetic Resonance (NMR) spectroscopy-based global profiling[49], together with Liquid Chromatography-Mass Spectrometry (LC-MS)-based targeted assays (e.g., SCFAs, bile acids, amino acids) will be applied according to in-house developed protocols[50-53].

Fiber Consumption Biomarkers

In our studies in Africa and the US, we have noted an association between fiber intake and the ratio of fecal SCFA to bile acids (BA). A ratio >10 is associated with low cancer risk, while <5 is associated with high risk ($p<0.0001$). We showed that increasing fiber 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[54, 55].

Gastric Motility/Intestinal Transit Time

A wireless capsule motility system (SmartPill Corporation, Buffalo, NY) 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 five 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 physiologic 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 (e.g., 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[56], epithelial apoptosis by cleaved caspase-3 staining[57], 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 defense, 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-fiber diets, which was reversed by resistant starch fiber supplementation (30 g/day) in human studies[9, 58, 59].

Diet Assessment/Compliance and Behavior Questionnaires

Two telephone 24-hour recall interviews will be conducted, one weekday and one weekend at each time point in Table 1. The interviews will be conducted using a multiple-pass interview with the Automated Self-Administered (ASA) 24-hour diet recall method[60]. We will use questionnaires designed to assess aspects of eating behavior 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 (DSQ)[35] and the 31-item Reasons Individuals Stop Eating Questionnaire (RISE-Q)[34] will be used at BL, 6, and 12 months. Visual analog scales (VAS)[36] 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 (6 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 (e.g., leisure, domestic, exercise) using the leisure time activity survey from the Cancer Prevention Study-3 (CPS-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 seven 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 fiber 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 organize, compile, and clean for statistical analysis. Data transfer will use secure internet protocols and will observe all IRB and 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 participant 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 programs[63]. Foodborne illness is a potential risk for participants; however, our strict inclusion and exclusion criteria and food safety protocol will help minimize the likelihood of occurrence. Also, the high-fiber diet may not be well tolerated in some participants (bloating, flatulence), however we will increase fiber 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 $\pm$ 0.9 kg. We expect to observe a larger difference at 6 mos. (e.g., 1.5 kg) and a 1.0 kg difference maintained at 12 mos. 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 ($\pm$ 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 enroll and randomize 70 participants, allowing for a dropout 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 fiber supplementation study with African Americans[12]. While a 20% difference between fiber 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 intent-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 (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 six 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 generalized linear models, Since the genes of interest have been specified in advance, control of the false discovery rate (FDR) 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 analyzed 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. Inter-sample distances (e.g. 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 modeling 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 analyze 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 (e.g., sex, age, BMI) in linear models. Based on their positive or negative correlation, ALR taxonomic values will be analyzed for correlation, to identify taxa with potential for cooperation (e.g., facilitation or syntrophy) or competition (e.g., displacement or predation). Distribution-based approaches, such as within-sample or alpha-diversity, analyze the sample profiles as probability distributions. As an index, taxonomic diversity will be analyzed 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 analysis (PCA) will be applied hierarchically to groups of related variables to identify principal components (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 analyzed[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 enrollment for consent and eligibility only. During the initial monitoring visit, 10% of total patient enrollment 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 westernized 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
[2]Department of Medicine, Division of Digestive Diseases, Emory University School of Medicine
[3,8,12,13]Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, University of Pittsburgh
[4]Department of Medicine, Division of Endocrinology, Metabolism and Lipids, Emory University School of Medicine
[5]Pulmonary, Allergy and Critical Care Medicine, Center for the Microbiome and Medicine, University of Pittsburg
[6]Department of Biostatistics and Bioinformatics, Department of Epidemiology, Rollins School of Public Health, Emory University
[7]Department of Nutritional Sciences, The Pennsylvania State University
[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
[10,11]Nutrition and Health Sciences, Rollins School of Public Health, Emory University

Acknowledgments 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. HH organized the original draft with contributions from all the co-authors. WDF designed the randomization 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 Intervention Development, Dissemination, and Implementation Developing Shared Resource of Winship Cancer Institute of Emory University and NIH/NCI under award number P30CA138292. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Georgia Clinical and Translational Science Alliance NIH/NCATS UL1 TR002378 award also contributed to the support of this research.

Competing interests No competing interests

Patient consent Not required

Ethics approval The study protocol was approved by the Emory Institutional Review Board (IRB approval number: 00000563).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Data will be shared in four ways: 1) preliminary and final results will be shared in the traditional conference presentation and manuscript formats; 2) results will be summarized and presented to the lay public via the NIH-supported CTSI listserv, news releases, and community lectures, as well as relevant entities at collaborator sites; 3) results will be summarized and presented via lectures and review articles to the academic community; 4) actual study data in raw and summary form, with protected health information removed, as well as access to the study database, will be available for reference of qualified professional colleagues and auditors up to 10 years after study publication. Data will be deposited in appropriate research data repositories (e.g., NIH Common Fund Metabolomics Workbench).

ORCID ID's Annette Wilson: 0000-0002-7426-1786
Helaina E. Huneault: 0000-0003-3866-2387
Jia V. Li: 0000-0002-5763-6670
Terryl J. Hartman: 0000-0001-5942-3161
Miriam R Pérez: 0000-0002-2432-4603
Ben Cousineau: 0000-0001-8732-2835
Barbara J. Rolls: 0000-0003-2374-1517
Blaine Loye Eberhart: 0000-0002-6239-053

References

1. Society AC. Key statistics for colorectal cancer 2022, January 12 [Available from: 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(8):794-8.
3. O'Keefe SJ. The association between fibre and hihg-income lifestyle-associated diseases: Burkitt's hypothesis revisited. *Lancet Gastro Hep*. 2019;4:in press.
4. Cummings JH, Engineer A. Denis Burkitt and the origins of the dietary fibre hypothesis. *Nutr Res Rev*. 2018;31(1):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(2):153-71.
6. O'Keefe SJ. Diet, microorganisms and their metabolites, and colon cancer. *Nat Rev Gastroenterol Hepatol*. 2016;13(12):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(6):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(9):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(8):786-95.
10. Bernstein C, Holubec H, Bhattacharyya AK, et al. Carcinogenicity of deoxycholate, a secondary bile acid. *Arch Toxicol*. 2011;85(8):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(4):315-22.
12. O'Keefe SJ, 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(6):445-50.
14. Hartman TJ, Zhang Z, Albert P, 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(2):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(1):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(1):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(6):513-20.
19. McCrory MA, Hamaker BR, Lovejoy JC, et al. Pulse consumption, satiety, and weight management. *Adv Nutr*. 2010;1(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(3):328-36.
21. Chaudhri O, Small C, Bloom S. Gastrointestinal hormones regulating appetite. *Philos Trans R Soc Lond B Biol Sci*. 2006;361(1471):1187-209.
22. Marinangeli CP, Jones PJ. 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 MJ, Santos A, Prada PO. Linking Gut Microbiota and Inflammation to Obesity and Insulin Resistance. *Physiology (Bethesda)*. 2016;31(4):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, Fernandez-Real JM. 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(5):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(1):85-94.
27. Ley RE, Backhed F, Turnbaugh P, et al. Obesity alters gut microbial ecology. *Proc Natl Acad Sci*. 2005;102(31):11070-5.
28. Ley RE, Turnbaugh PJ, Klein S, et al. Microbial ecology: Human gut microbes associated with obesity. *Nature*. 2006;444(7122):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(2):241-7.
30. O'Keefe SDC, D Mahmoud, N, Sepulveda AR, and Manafe M. Why Do African Americans Get More Colon Cancer than Native Africans? *J Nutr*. 2007;137(1):175S-82.
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(7):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. *J Am Diet Assoc*. 2008;108(5):883-8.
33. Hall KD, Kahan S. Maintenance of Lost Weight and Long-Term Management of Obesity. *Med Clin North Am*. 2018;102(1):183-97.
34. 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;161:105127.
35. James BL, Loken E, Roe LS, et al. Validation of the Diet Satisfaction Questionnaire: a new measure of satisfaction with diets for weight management. *Obesity science & practice*. 2018;4(6):506-14.
36. 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(1):38-48.
37. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey (NHANES) Anthropometry Procedures Manual. Atlanta, GA; 2016.
38. 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(7):2099-106.
39. 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 7α-dehydroxylating bacteria in human feces. *Clin Chim Acta*. 2003;331(1-2):127-34.
40. 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(7423):254-8.

41. Kostic AD, Chun E, Robertson L, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe*. 2013;14(2):207-15.
42. Mima K, Sukawa Y, Nishihara R, et al. Fusobacterium nucleatum and T Cells in Colorectal Carcinoma. *JAMA Oncol*. 2015;1(5):653-61.
43. Devkota S, Wang Y, Musch MW, et al. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. *Nature*. 2012;487(7405):104-8.
44. Attene-Ramos MS, Wagner ED, Plewa MJ, et al. Evidence That Hydrogen Sulfide Is a Genotoxic Agent. *Mol Cancer Res*. 2006;4(1):9-14.
45. 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(17):5112-20.
46. 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.
47. 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(Database issue):D141-5.
48. 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(Database issue):D590-6.
49. 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(19):9887-94.
50. Lozupone C, Hamady M, Knight R. UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics*. 2006;7:371.
51. 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(5):1576-85.
52. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8(4):e61217.
53. 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(19):9662-70.
54. 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(1):111-20.
55. 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(11):869-80.
56. 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(6):760-75.
57. 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(4):449-54.
58. Cummins JM, He Y, Leary RJ, et al. The colorectal microRNAome. *Proc Natl Acad Sci U S A*. 2006;103(10):3687-92.
59. 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(6):459-74.

60. 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(8):1134-7.
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(3):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(2):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(5):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(4):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(434):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* 1995;57(1):298-310.
69. Matthews JNS, Atlman DG, Campbell JJ, et al. Analysis of serial measurements in medical research. *Br Med J*. 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(1):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(1):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.

Data Collection – Supplementary Materials:

1) **Blood collection:** Approximately 50 ml of blood will be collected from the subjects at baseline, 6 and 12 months, and at extended follow-up. Samples will be collected after an overnight fast and in the early morning. Serum and plasma samples will be aliquoted in smaller vials, frozen at –80°C, and stored for later analysis of insulin, glucose, and CRP using standard techniques. Medication use will be assessed at each collection.

2) **Insulin and Insulin resistance measures:** As mentioned above, fasting blood samples will be collected, aliquoted and stored frozen at -80°C for analysis at study completion. Insulin resistance will be evaluated by the Homeostasis Assessment Model from fasting insulin and glucose [= fasting insulin (μU/mL) × fasting glucose (mmol/L)/22.5]; values > 2.61 will be considered insulin resistant [1]. Samples will be analyzed under the direction of Dr. Ngoc-Anh Le, Biomarker Core Laboratory at the Atlanta VA Medical Center. Glucose will be determined by colorimetric methods (Sekisui Diagnostics, Exton, PA), plasma insulin levels assessed using the immunoturbidometric method (Sekisui), and high sensitivity CRP analyzed via sandwich enzyme immunoassay (ALPCO). Samples will be grouped in random order for analysis, and a 10% blind quality control included. The CV for these analyses in the Le lab ranges between 3.9%-6.1%.

3) **Urine Collection:** 50 ml urine samples will be collected at baseline, 6 and 12 months and at extended follow-up in plastic containers by clean-catch technique, transported, and stored in the same way as fecal samples. Samples will be aliquoted into five containers/time point, coded by GCRC laboratory staff, and held at –80°C for future metabolome analysis led by Dr. O'Keefe [2].

4) **Fecal Collection:** Stool samples will be collected during the study. Subjects will be instructed in the use of a plastic device to cover the toilet seat and collect the stool. Two separate ~5 g samples will be taken for fecal microbiome analyses (mechanistic studies) and two additional samples for fecal SCFA and bile acid analyses (markers of compliance with HLD) at the time points specified in Table 1. All samples will be transported to the laboratory, coded by the research assistant, and held at –80° C for future DNA extraction and microbiome analysis by targeted and global approaches led by the O'Keefe laboratory.

5) Body composition: Body composition will be assessed using dual energy X-ray absorptiometry (DXA)[3] in the GCRC at Emory University. This method uses a whole-body scanner to measure total body composition and fat content with a high degree of precision. It is safe and noninvasive with little burden to the individual. Data from the DXA scans will be used to assess longitudinal changes in body fat that accompany weight loss. This data will also allow examination of changes in fat distribution at defined regions in the body. New software allows for the estimation of visceral fat. Women who could potentially become pregnant will be given a pregnancy test prior to a DXA.

References:

1. Monzillo LU, Hamdy O. Evaluation of insulin sensitivity in clinical practice and in research settings. *Nutr Rev*. 2003;61(12):397-412.
2. O'Keefe SJ, Li JV, Lahti L, et al. Fat, fibre and cancer risk in African Americans and rural Africans. *Nat Commun*. 2015;6:6342.
3. Lunar GH. X-ray bone densitometer with enCORE v17 software—user manual. *Madison: GE Healthcare Lunar*. 2016.

Meal Plan A – Healthy American Control group meal plan

Emory GCRC Bionutrition

Meal	1	2	3	4	5	6	7
Breakfast							
Lunch	Lemon Chicken Potato Casserole	Quinoa Tofu Pesto	Turkey Burger on a roll (side of: lettuce, onion, catsup)	South of the Border Chicken Stew	Bulgur-Beef Meatballs with Spaghetti & Sauce	Lemon Chicken w/ Bulgur	Spinach-Cheese Noodles
Sides	*Non-Starchy Vegetables (i.e. Roasted Brussel Sprouts), Fruit (i.e. Pear)*	*Non-Starchy Vegetable (i.e. Steamed Asparagus), Fruit (i.e. Orange)*	*Non-Starchy Vegetable (i.e. Sliced Tomatoes), Fat (i.e. Avocado)*	*Starch (i.e. Whole Grain Sliced Bread), Fat (i.e. Avocado)*	*Vegetable (i.e. Salad)*	*Non-Starchy Vegetable (i.e. Green Beans), Fruit (i.e. Nectarine)*	*Protein (i.e. Grilled Salmon), Non-Starchy Vegetable + Combination Food [Carbohydrate & Fat] (i.e. Raw Sweet Peppers & Hummus)*
Dinner	Pasta & Peanut Salad	Stovetop BBQ Chicken with basmati rice & raw spinach	Spaghetti with Turkey Meat Sauce	Turkey Meatloaf	Spiced Chicken & Rice	Potato Green Bean Bake	Beef & Broccoli Stir Fry with Basmati Rice
Sides	*Starch (i.e. Pretzels), Fruit (i.e. Tangerine)*	*Starchy Vegetable (i.e. Corn on the Cob)*	*Starchy Vegetable (i.e. Mixed Vegetable, frozen), Fruit (i.e. Pear)*	*Starch Vegetable (i.e. Baked Potato), Vegetable (i.e. Salad)*	*Non-Starchy Vegetable (i.e. Steamed Broccoli), Fat (i.e. Almonds)*	*Vegetable (i.e. Salad), Non-Starchy Vegetable + Fat (i.e. Celery & Peanut Butter)*	*Fruit (i.e. Seedless Grapes)*

Meal Plan B – Legume Intervention group meal plan Emory GCRC Bionutrition

Meal	1	2	3	4	5	6	7
Breakfast							
Lunch	Chickpea Salad Sandwich (inc. whole grain bread, lettuce)	Lentil Tomato Salad	Black Bean Salad	Black-Eyed Pea Curry w/ Basmati Rice	Chana Massala (flavorful chickpea, tomato & spinach stew) w/ Basmati Rice	Falafel (baked chickpea patties) [served with whole wheat pita, lettuce, tomato, onion, tahini]	Southwest Bean Soup
Sides	*Non-Starchy Vegetable (i.e. Tomato Slices), Fruit (i.e. Apple)*	*Starchy Vegetable (i.e. Baked Sweet Potato), Fruit (i.e. Apple)*	*Combination Food (carbohydrate, fat) [i.e. Tortilla Chips], Non-Starchy Vegetables (i.e. Cherry Tomatoes)*	*Fruit (i.e. Orange)*	*Non-Starchy Vegetable + Combination Food (Carbohydrate & Fat) [i.e. Baby Carrots & Hummus]*	*Non-Starchy Vegetable (i.e. Sliced Cucumber)*	*Vegetable (i.e. Salad), Fruit (i.e. Apple)*
Dinner	Pineapple Lentils w/ Basmati Rice	Quinoa Chili	Bean & Chicken Cassoulet (casserole)	Caldo Verde (Potato, Kale & White Bean Soup)	White Bean Onion Stew	African Peanut Stew	Tamale Pie
Sides	*Vegetable (i.e. Salad)*	*Fat (i.e. Avocado), Fruit (i.e. Peach)*	*Vegetable (i.e. Salad), Non-Starchy Vegetable (i.e. Steamed Broccoli)*	*Starch (i.e. Whole Grain Toast), Medium-Fat Protein (i.e. Poached or Hard-Boiled Egg)*	*Vegetable (i.e. Salad), Medium-Fat Protein (i.e. Hard-Boiled Egg)*	*Non-Starchy Vegetable (i.e. Celery Sticks), Fruit (i.e. Seedless Grapes)*	*Non-Starchy Vegetable (i.e. Baby Carrots), Fruit (i.e. Tangerine)*

SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents*

Section/item	Item No	Description
Administrative information		
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym Page 1
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry Page 2
	2b	All items from the World Health Organization Trial Registration Data Set N/A
Protocol version	3	Date and version identifier Page 2
Funding	4	Sources and types of financial, material, and other support Page 13
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors Page 13
	5b	Name and contact information for the trial sponsor Page 13
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities Page 13
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee) Page 12
Introduction		

Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention Page 4,10
	6b	Explanation for choice of comparators Pages 5-10
Objectives	7	Specific objectives or hypotheses Page 4,5
Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory) Page 4-7
Methods: Participants, interventions, and outcomes		
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained Page 5, 6
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists) Page 5
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered Page 5-7
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease) Page 10
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests) Page 10
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial Page 6

Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended Page 4-9
Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure) Page 7
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations Page 10,11
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size Page 5

Methods: Assignment of interventions (for controlled trials)

Allocation:

Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions Page 6
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned Page 6
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions Page 6
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how Page 5,6

	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial Page 5
Methods: Data collection, management, and analysis		
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol Pages 7-10
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols Page 10
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol Page 10
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol Page 10-12
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses) Page 10-12
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation) Page 10-12
Methods: Monitoring		
Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed Page 12

	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial Page 10,11
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct Page 10
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor Page 10,12

Ethics and dissemination

Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval Page 12,13
Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators) Page 12,13
Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32) Page 5
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable N/A
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial Page 5
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site Page 13
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators Page 13
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation N/A

Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions Page 12
	31b	Authorship eligibility guidelines and any intended use of professional writers Page 13
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code N/A
Appendices		
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates Supplementary Materials
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable Page 7

*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group under the Creative Commons "Attribution-NonCommercial-NoDerivs 3.0 Unported" license.