PERSPECTIVE

Perspective: Design and Conduct of Human Nutrition Randomized Controlled Trials

Alice H Lichtenstein,[1] Kristina Petersen,[2] Kathryn Barger,[3] Karen E Hansen,[4] Cheryl AM Anderson,[5] David J Baer,[6] Johanna W Lampe,[7] Helen Rasmussen,[8] and Nirupa R Matthan[8]

[1]Cardiovascular Nutrition Laboratory, Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, USA; [2]Pennsylvania State University, University Park, PA, USA; [3]Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, USA; [4]School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA; [5]Division of Preventive Medicine, Department of Family Medicine of Public Health, University of California, San Diego, La Jolla, CA, USA; [6]Food Components and Health Laboratory, Beltsville Human Nutrition Research Center, USDA Agricultural Research Service, Beltsville, MD, USA; [7]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; and [8]Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA

ABSTRACT

In the field of human nutrition, randomized controlled trials (RCTs) are considered the gold standard for establishing causal relations between exposure to nutrients, foods, or dietary patterns and prespecified outcome measures, such as body composition, biomarkers, or event rates. Evidence-based dietary guidance is frequently derived from systematic reviews and meta-analyses of these RCTs. Each decision made during the design and conduct of human nutrition RCTs will affect the utility and generalizability of the study results. Within the context of limited resources, the goal is to maximize the generalizability of the findings while producing the highest quality data and maintaining the highest levels of ethics and scientific integrity. The aim of this document is to discuss critical aspects of conducting human nutrition RCTs, including considerations for study design (parallel, crossover, factorial, cluster), institutional ethics approval (institutional review boards), recruitment and screening, intervention implementation, adherence and retention assessment, and statistical analyses considerations. Additional topics include distinguishing between efficacy and effectiveness, defining the research question(s), monitoring biomarker and outcome measures, and collecting and archiving data. Addressed are specific aspects of planning and conducting human nutrition RCTs, including types of interventions, inclusion/exclusion criteria, participant burden, randomization and blinding, trial initiation and monitoring, and the analysis plan. *Adv Nutr* 2021;12:4–20.

Keywords: nutrition, diet, design, randomized controlled trials, recruitment, screening, adherence, retention, scientific integrity

Introduction

In the field of human nutrition, randomized controlled trials (RCTs) are considered the gold standard for establishing causal relations between exposure to nutrients, foods, or dietary patterns and prespecified outcome measures, such as body composition, biomarkers, or event rates. Evidence-based dietary guidance is frequently derived from systematic reviews and meta-analyses of these RCTs (1, 2). However, efforts to prepare evidence-based dietary guidance are often hampered by the lack of sufficiently large databases from which to formulate the guidance. In some cases, a large body of evidence is available, but its robustness is limited by a preponderance of low-quality studies relative to the question of interest, either due to methodological limitations or failure to document critical aspects of the intervention. The latter issue is more common for work published prior to the introduction of standardized reporting guidelines—the Consolidated Standards of Reporting Trials (CONSORT). CONSORT (3), now adopted as a standard by the International Committee of Medical Journal Editors and many peer-reviewed journals, established a minimum set of criteria for reporting randomized trials (**Table 1**). Studies reported according to the CONSORT guidelines are more likely to be included in the formulation of evidence-based guidance. A complete set of additional reporting guidelines for different study designs can be found at the EQUATOR (Enhancing the Quality and Transparency of Health Research) network website, along with the criteria used to rank studies (2). The aim of this report is to discuss critical aspects of conducting a human nutrition RCT, including considerations for study design, institutional ethics approval, recruitment and screening, intervention implementation, adherence and retention assessment, and sample size considerations.

 Adv Nutr 2021;12:4–20; doi: https://doi.org/10.1093/advances/nmaa109.

Designing human nutrition RCT protocols

Human nutrition RCTs can be designed to generate biomarker and/or outcome data, identify mechanisms, or develop and validate new technologies or methods. Each decision made during the design of a human nutrition RCT will affect the utility or the generalizability of the study results. Within the context of limited resources, the goal is to maximize the generalizability of the findings while producing the highest quality data and maintaining the highest levels of ethics and scientific integrity.

Research question and specific aims

The initial step in designing a research protocol is to identify ≥1 key research questions. One approach to accomplish this is summarized by the FINER (feasible, interesting, novel, ethical, and relevant) criteria (**Table 2**) (4). In human nutrition RCTs, there are unique challenges to each of these criteria. For example, feasibility may be limited by study facility capacity, amount of test material available (e.g., supplement), and fiscal and personnel resources; interest, novelty, and relevance can be dependent on timing of the work relative to other available data; and ethical standards need to be considered within the context of safety and degree of participant burden.

The actual research questions addressed are defined by the specific aims of the study. Critical components of a well-structured and precise research question, enumerated by the PICO acronym, include population (P), intervention (I), comparator (C), and outcome (O) (**Table 3**). When there are multiple treatment groups and outcome measures, planned comparisons must be specified a priori to account for the number of statistical tests performed. Such considerations control for the type I error rate (discussed below) and affect sample size calculations. Specific aims (primary and secondary) and exploratory aims, as well as hypotheses, are directly derived from the research question(s).

Study designs

The choice of study design is driven by the specific aims and unique characteristics of the intervention. Commonly utilized designs for human nutrition RCTs include parallel, crossover, factorial, and cluster. In addition, there are more advanced designs that can be used (5). Each study design has strengths and weaknesses, examples of which are shown in **Table 4**. In some cases, the choice of study design is dictated by feasibility (e.g., facilities, staffing, availability of target population). Regardless of study design, the duration of the intervention is determined by the predicted period necessary for the outcome(s) of interest to reach a biologically relevant change. For continuously measured outcomes, it is important to consider how the outcome will change during the intervention period, whether it would be predicted to reach a plateau value (e.g., RBC fatty acid profiles, blood lipid and lipoprotein concentrations) or to have a linear trajectory (e.g., body weight, blood pressure).

Parallel study design.

In a parallel study design, each participant is randomly assigned to a single study intervention (**Figure 1**A). Comparisons between groups (assuming 2 groups, or among groups if ≥3) are based on between-group/among-subject variation. Treatment groups can be balanced (equal number of participants in each group, similar demographic characteristics) or unbalanced. The parallel study design is simple and has the advantage that multiple treatments can be tested simultaneously, resulting in a shorter study duration than a crossover study design (6). Moreover, extraneous differences, such as unanticipated changes in the composition of study foods or participants' seasonal exposure to UV light, potentially affecting serum 25-hydroxyvitamin D_3 concentrations, are minimized. Likewise, issues related to potential treatment carryover effects do not arise. A limitation is that a parallel study design requires a larger sample size than a crossover study design due to participant characteristic variability between/among groups.

Randomized crossover study design.

In a crossover study design, each participant receives a set of treatments in a random order or fixed sequence (Figure 1B). The crossover design can be used when the treatment effect is considered reversible—that is, after an adequate intervention period followed by a washout period, the change in the outcome measure is predicted to have no carryover effect from 1 intervention period to another (e.g., blood pressure, plasma cholesterol concentrations). Since each participant serves as his/her own control, crossover study designs are unsuitable when outcomes are unlikely to reverse in the short-term (e.g., weight loss, correcting nutrient inadequacy) or have long carryover effects (e.g., change in RBC cell fatty acid profile or hepatic fat content) (7). Treatment effects are based on within-participant variation and these designs have higher statistical power relative to the sample size than non–crossover study designs. However, the gain in statistical power is linked to higher participant burden associated with a longer total intervention period, and consequently increased risk of dropouts and complexity of the statistical analyses (8). In the analyses, potential carryover effects should be assessed

This project was funded by National Institutes of Health Clinical and Translational Science Awards to Tufts Clinical and Translational Science Institute (UL1TR002544), Indiana Clinical and Translational Sciences Institute (UL1TR002529), and Penn State Clinical and Translational Science Institute (UL1TR002014).
Author disclosures: The authors report no conflicts of interest.
Perspective articles allow authors to take a position on a topic of current major importance or controversy in the field of nutrition. As such, these articles could include statements based on author opinions or point of view. Opinions expressed in Perspective articles are those of the author and are not attributable to the funder(s) or the sponsor(s) or the publisher, Editor, or Editorial Board of *Advances in Nutrition*. Individuals with different positions on the topic of a Perspective are invited to submit their comments in the form of a Perspectives article or in a Letter to the Editor.
The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
Supplemental Tables 1–3 and Supplemental Figure 1 are available from the "Supplementary data" link in the online posting of the article and from the same link in the online table of contents at https://academic.oup.com/advances.
Address correspondence to AHL (email: alice.lichtenstein@tufts.edu).
Abbreviations used: CONSORT, Consolidated Standards of Reporting Trials; DASH, Dietary Approaches to Stop Hypertension; FWER, family-wise error rate; IRB, institutional review board; RCT, randomized controlled trial; SOP, standard operating procedure.

TABLE 1 CONSORT guidelines[1]

	Guidelines
Title	Identification as a randomized trial in the title
Abstract	Structured summary of trial design, methods, results, and conclusions
Introduction	
Background	Scientific background and explanation of rationale
Objective	Specific objectives or hypotheses
Methods	
Trial design	Description of design (e.g., parallel, factorial) including allocation ratio
Changes to trial design	Important changes to methods after trial commencement (such as eligibility criteria), with reasons
Participants	Eligibility criteria for participants
Study setting	Settings and locations where the data were collected
Interventions	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered
Outcomes	Completely defined prespecified primary, secondary, and exploratory specific aims, including how and when they were assessed
Changes to outcomes	Any changes to trial outcomes after the trial commenced, with reasons
Sample size	How sample size was determined
Interim analyses and stopping guidelines	When applicable, explanation of any interim analyses and stopping guidelines
Randomization: sequence generation	Method used to generate the random allocation sequence
Randomization: type	Type of randomization; details of any restriction (such as blocking and block size)
Randomization: allocation concealment mechanism	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned
Randomization: implementation	Who generated the allocation sequence, who enrolled participants, and who assigned participants to interventions
Blinding	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how
Similarity of interventions	If relevant, description of the similarity of interventions
Additional analyses	Methods for additional analyses, such as subgroup analyses and adjusted analyses
Results	
Participant flow	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome
Losses and exclusions	For each group, losses and exclusions after randomization, together with reasons
Reason for stopping trial	Why the trial ended or was stopped
Baseline data	A table showing baseline demographic and clinical characteristics for each group
Numbers analyzed	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups
Outcomes and estimation	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% CI)
Binary outcomes	For binary outcomes, presentation of both absolute and relative effect sizes is recommended
Ancillary analyses	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory
Harms	Important harms or unintended effects in each group
Discussion	
Limitations	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses
Generalizability	Generalizability (external validity, applicability) of the trial findings
Interpretation	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence
Other information	
Registration	Registration number and name of trial registry
Protocol	Where the full trial protocol can be accessed, if available
Funding	Sources of funding and other support (such as supply of drugs), role of funders

[1] Adapted from reference 3. CONSORT, Consolidated Standards of Reporting Trials.

by testing for treatment-by-period interaction or evaluating baseline measurements collected at the start of each period.

Factorial study design.

In a factorial study design, combinations of ≥2 interventions are randomly assigned to each participant, allowing evaluation of both main effects and interactions of multiple treatments at the same time within a single study (Figure 1C). The factorial study design is efficient since balanced group sizes in each combination of the interventions provide optimal statistical power for both estimating interactions and performing subgroup analyses. A 2 × 2 factorial design is the most common approach, involving 2 interventions at 2 levels each. Factorial designs are used when the interventions can be assigned independently and it is of interest to

TABLE 2 FINER—criteria for defining research questions[1]

	Criteria
Feasibility	Scope within resources
Interesting	Topic of contemporary importance
Novel	Confirms, refutes, or extends prior work
Ethics	Conforms to current guidance
Relevance	Advances prevailing scientific knowledge

[1] Adapted from reference 4. FINER, feasible, interesting, novel, ethical, and relevant.

examine additive, synergistic, or antagonistic effects. If the interaction effect is the primary outcome, sample size must be calculated carefully, as sample size requirements are ~4 times larger than a parallel study evaluating a single intervention (**Box 1**).

Cluster study design.

In a cluster study design, interventions are randomly assigned to entire groups rather than individuals (Figure 1D). Cluster study designs can be an efficient way to administer an intervention in a population with natural groupings, such as households, schools, classrooms, or community centers. Sample size computations for these trials must account for intracluster correlation, for which there may be limited preliminary estimates. Randomization to clusters is advantageous in settings where it would be otherwise difficult to assign different treatments to participants within the same cluster (e.g., different dietary patterns for individuals living together in the same household, eating in the same school cafeteria). Both the number of clusters and number of individuals within a cluster must be selected to maximize statistical power and facilitate implementation of the study.

Efficacy and effectiveness

A major distinction in human nutrition RCTs is efficacy versus effectiveness. Efficacy refers to the consequence of an intervention under ideal conditions. Typically, in efficacy studies, the intervention delivery will involve complete provision of foods, beverages, and/or a nutrient formulation in either an inpatient setting (e.g., conducted in a metabolic ward, hospital) or to free-living participants (e.g., distributed through a metabolic research unit, test kitchen). Participants are carefully monitored (inpatient setting) or may eat ≥1 meals/d under supervision and the balance on their own (e.g., free-living setting). Available resources, time, and logistics usually narrow the size and breadth of participant characteristics in an efficacy study. Examples of efficacy studies include the Dietary Effects on Lipoproteins and Thrombogenic Activity Trial (9, 10), Dietary Approaches to Stop Hypertension (DASH and DASH-Sodium) trials, and the OmniHeart and OmniCarb trials (11, 12, 13).

Box 1.
An example of study design considerations in a human nutrition RCT: VITAL stud (14, 15)

A 2 × 2 factorial design to look at the effect of vitamin D and omega-3 fatty acids for cancer and cardiovascular disease (CVD) prevention.

The Vitamin D and Omega-3 Trial (VITAL) randomly assigned individuals to vitamin D supplementation, omega-3 fatty acid supplementation, both agents, or both placebos for 5 y. Over 25,000 older adults in the United States were enrolled in the study and followed to evaluate cancer and CVD cases. The study was a double-blind 2 × 2 factorial trial. Treatment effects were considered for each intervention agent separately and study results were published in 2 articles.

How aims were specified:

The primary aim was to test for reduction in risk for total cancer and major CVD events (as a composite endpoint). The secondary aims were to test for treatment effects in specific cancers, total cancer mortality, and individual CVD endpoints. Tertiary aims were specified to describe other exploratory outcomes.

How subjects were randomized:

Individuals were randomly assigned to each treatment in blocks of 8 individuals. Randomization was stratified by 5-y age groups to ensure balance and increase statistical efficiency.

Description of intervention and placebos:

The active agents were vitamin D_3 2000 IU/d or placebo, and omega-3 fatty acid (EPA + DHA, in the ratio 1.3 to 1) 1 g/d. Placebo pills were used for each active agent, such that participants in all groups took 2 pills each day.

How statistical power was calculated:

Statistical power was calculated for main effects from the factorial design. Factorial designs enable estimation of interaction effects; however, in VITAL, testing interactions between the interventional agents was specified as an exploratory outcome. The trial was designed to have a greater than 85% power to detect observed HRs of 0.85 and 0.80 for the primary endpoints of cancer and CVD, respectively. Calculations were based on a 2-sided log-rank test with a significance level of 0.05 for each outcome.

How multiplicity due to testing multiple outcomes was addressed:

Each outcome was considered with a type I error of 0.05 and, correspondingly, statistical power calculations were based on 0.05 used for each outcome.

Other design considerations:

The study plan was to include at least 5000 Black participants. A 3-mo placebo run-in period was used to select participants with anticipated high adherence to the study protocol and long-term follow-up.

If an efficacy study yields clinically significant outcomes, the next step would be to conduct an effectiveness trial. An effectiveness trial, also referred to as a pragmatic trial, is designed to reflect a "real world" situation, with implementation of an intervention under less-controlled conditions. It typically has a lower participant burden and involves a larger number of participants. In this scenario, investigators instruct participants, either on an individual basis, in a group setting, or by a Web-based meeting, on how to modify their diet (e.g., instruction on purchase,

TABLE 3 PICO—formulating a research question

	Components of PICO questions	
P	Population	Study population characteristics, defined on the basis of inclusion/exclusion criteria (e.g., age, sex, BMI, genotype, risk factor ranges)
I	Intervention	Supplement, food(s)/beverage(s), dietary pattern
C	Comparator (comparison)	Placebo or different supplement, food(s)/beverage(s), dietary pattern
O	Outcome	Biomarkers, clinical measures, event rates

preparation, and/or substitution of specific foods and beverages), with or without the provision of unique study items. Examples of effectiveness studies include the PREMIER study and Women's Health Initiative Intervention Study (16, 17).

Biomarker and outcome measures

Both safety- and intervention-related biomarker and outcome measures are determined by the study-specific aims. For the latter, due to the time delay between biological changes and hard clinical outcomes, human nutrition RCTs often use intermediate biomarkers of disease risk (i.e., risk factors or risk predictors), such as measures of inflammation and blood lipids, or more recently, whole metabolomes, to evaluate the effect of an intervention. In many cases, evidence supports a causal relation but other mitigating factors (e.g., body weight, body fat distribution, physical activity, exposure to tobacco products) may alter the relations; hence, biomarker data need to be interpreted with caution. Notwithstanding these issues, statistical power calculations in human nutrition intervention trials are typically based on predicted changes in biomarkers rather than event rates. Examples of prespecified safety measures are included in **Supplemental Table 1**.

A large number of primary and secondary outcome measures, as dictated a priori by the study-specific aims, has the advantage of increasing the scope of the hypotheses tested (see **Supplemental Table 2** for an example of a human nutrition RCT procedure schedule and measures). However, disadvantages of a large number of outcome measures include an increase in participant burden due to the number of biological samples required and/or number of study visits and amount of time (e.g., questionnaires, anthropometric measures), as well as necessary resources and diminished statistical power due to corrections for multiple comparisons.

Data collection

Standard operating procedures.

Standard operating procedures (SOPs) should be developed to ensure that all data are collected, processed, and stored optimally and consistently. SOPs promote strict adherence to the study protocol and consistency in sample processing among study participants, as well as research groups if the intervention is multisite.

Questionnaire data.

Questionnaires are an integral component of data collection during human nutrition RCTs. For human nutrition RCTs, questionnaire data are typically collected using self-administered paper or electronic tools, interviewer-administered phone or in-person protocols, or a combination of both. Baseline questionnaires permit accurate characterization of potential participants, including such factors as demographic characteristics, comorbid diseases, anthropometric characteristics, medication use, tobacco exposure, alcohol intake, habitual dietary intake, and physical activity patterns. Evaluation of habitual diet and physical activity patterns is particularly important for estimating energy needs. Demographic and health history data, either collected by self-report or direct measurement, may be used in block randomization or as covariates in statistical modeling. Questionnaires administered during the intervention period are critical for capturing adherence and tolerance to the intervention, as well as additional information that may affect study outcomes, such as minor illness, unanticipated travel, change in medication use, over-the-counter medication use, and phase of menstrual cycle.

Anthropometric characteristics and biomarker data.

It is critical to identify appropriate anthropometric characteristics (e.g., height, weight, BMI, and hip and waist circumferences), biomarkers (e.g., serum cholesterol, glucose, insulin concentrations) and outcomes (blood pressure, glomerular filtration rate, flow-mediated dilation) prior to the start of a human nutrition RCT, as well as standardized equipment and collection, aliquoting, and storage conditions, respectively. This type of information should be included in the SOPs. Examples of issues to specify include blood fraction [serum, plasma (specific anticoagulant), buffy coat, RBCs], storage conditions (temperature, light, and, if the analyte is pH sensitive, how to adjust to the appropriate pH), or if the analyte is susceptible to oxidation, determine the antioxidant to add and/or whether to replace the air with nitrogen (flush with nitrogen) prior to freezing. Additional examples of potential factors that should be specified in SOPs include processing and storage temperature, light exposure, and sample volumes for specimen aliquots. Sample labeling schemes should not contain ambiguous information or personal identifiers. A system should be developed for monitoring sample usage, storage, and availability. Written SOPs should be developed for all assays [additional detail is provided in reference (18)]. All study personnel should

TABLE 4 Common human nutrition RCT study designs[1]

Study design	Strengths	Weaknesses	Other considerations	Examples of nutrition.
Parallel	Multiple interventions assessed simultaneously; hence, relatively short duration due to single exposure per participant and no washout period No carryover effects Minimizes unanticipated variability during study period (e.g., change in composition of study food, duration of sunlight exposure	Requires larger sample size than crossover design to achieve similar statistical power	Risk unbalanced characteristics participant groups	POUNDSLOST (19) DIETFITS (20)
Crossover	Requires smaller sample size than parallel design to achieve similar statistical power	Higher participant burden than crossover design due to multiple intervention and washout periods Long duration increases risk of dropouts Introduces potential carryover effects	Adequacy of washout period length difficult to estimate a priori	DELTA (9, 10) DASH (11) OmniHeart (12)
Factorial	Resource and time efficient, simultaneously evaluate multiple interventions independently and combined	Requires relatively large sample size	Potential confounding between/among interventions should be factored into sample size calculations and statistical analyses	VITAL (15, 14) WHI (16) AREDS AMD (21)
Cluster	Natural groupings of participants facilitate recruitment and delivering intervention (e.g., households, schools, community centers)	Uneven cluster size complicates statistical analysis Necessity to match clusters may limit potential participant pool	Need to balance sample size to account for both number of clusters and participant number within each cluster	SSaSS (22) SMART (23) Improve health choices (24)

[1] DASH, Dietary Approaches to Stop Hypertension; RCT, randomized controlled trial; SSaSS, Salt Substitute and Stroke Study; VITAL, Vitamin D and Omega-3 Trial; WHI, Women's Health Initiative.

review these procedures and be trained appropriately, and these procedures should be readily available to all study personnel to ensure adherence.

Outcome data.
In some human nutrition RCTs, "hard endpoints" or disease outcomes, such as incident events (e.g., stroke, myocardial infarction) or disease status (e.g., carotid artery thickness, arterial calcium score) are prespecified. Human nutrition RCTs are less likely to use hard clinical outcomes than in other types of studies because, in addition to the long lead times for natural disease progression, there are other challenges, such as sustaining adherence to a dietary modification (avoiding recidivism). Hence, long-term human nutrition studies are most often observational in nature.

Conducting Human Nutrition RCTs

Alignment of diet and intervention with specific aim(s)

A key component of human nutrition RCT intervention design construction is ensuring that the nutrition variable of interest and the comparator(s) are consistent with the specific aims of the study. Nutrition research poses unique challenges in this regard because a modification of an energy-containing component (e.g., test food or beverage addition, change in absolute amount of saturated fat) triggers compensatory changes in the composition of the diet that in itself may alter results and interpretation of the findings. Compensatory changes might include a change in total energy intake, relative proportion of macronutrients, or type (quality) of macronutrient. One way to control for this effect is to identify a suitable comparator for a test food or beverage, so that a consistent iso-energy exchange occurs: for example, comparing the effects of refined flour with whole-wheat flour, or olive oil with soybean oil, on systematic inflammation (25). Or, to assess the effect of saturated fat on LDL-cholesterol concentrations, prespecify whether the iso-energy exchange will be monounsaturated fat, polyunsaturated fat, refined carbohydrate, or unrefined carbohydrate (26, 27). In both examples, the comparator should be driven by the study-specific aims. If a replacement energy source is not specified in the protocol (e.g., increase vegetable sources of protein), attention should be paid to capturing daily food intake to determine whether and if so, what, compensations were made to total intake so these data can be factored into the interpretation of the biomarkers and body-weight data. Thus, careful alignment of the intervention with the study-specific aims is critical.

Types of interventions

Nutritional supplements.
Studies of nutritional supplements or single-nutrient studies are more amenable to a double-blinded study design (see section entitled "Intervention blinding") because a placebo can usually be designed to match the supplement. The baseline nutrient status of participants should be assessed using a valid and reliable biomarker of intake, and then accounted for in both the study design and data interpretation (28). For example, serum 25-hydroxyvitamin D concentration is used as a biomarker for vitamin D status and plasma vitamin B-12 concentration is used for vitamin B-12 status. Some nutrients, for example, sodium and calcium, have no commonly accepted biochemical indicators. Assessment of nutrient status must be estimated using multiple 24-h urine collections for the former and self-reported intake for the latter (29). Other considerations for nutrient supplementation trials include dosage, bioavailability of test formulation, stability under different storage conditions, timing of supplementation relative to food intake, and co-administration with other foods, nutrients, or bioactive compounds (e.g., iron and phytates) that may impact bioavailability or efficacy (28). The composition and stability of the supplement should be verified by laboratory analysis.

Single food/beverage or dietary component item.
For interventions using a single food/beverage item, detailed instructions should be provided to participants as to when and with what to consume the test item, with what (or without what) other foods and beverages to consume the test item, and appropriate exchange items to minimize introducing heterogeneity. In the absence of instruction, participants might add the test food item to their habitual diet rather than replacing with another item, resulting in weight gain (30). Alternatively, providing a food without instructions on how to incorporate it into a habitual diet will allow an assessment of common dietary recommendations involving a single food or food group recommendations [e.g., eat 2.5 cups (∼380 g) of vegetables/d]. Provision of detailed instructions or nutrition counseling will increase the likelihood that the participants will incorporate the food or dietary component in the way intended by the study design and minimize risk of unintended changes that may confound the results.

Controlled-feeding trial.
In a controlled-feeding study, all foods and beverages are provided for either consumption on or off site. Design aspects to consider include diet composition, nutrient adequacy, menu rotation (to avoid boredom and nonadherence), non–study food/beverage lists consistent with the study protocol (e.g., low-energy beverages, herbs, and spices), food procurement and storage, and estimation of energy requirements consistent with protocol-specific body-weight goals. Depending on the study design, a run-in period, adherence breaks, or a washout period may be required. Chemical analysis should be completed prior to the start of the study to ensure the calculated values reflect the actual target composition consistent with the study aims (31). Periodic chemical analyses of the intervention food(s)/diet(s) should be scheduled to monitor potential drift in the diet provided to the study participants. Collecting analytical data on the nutrient composition of study diets will aid in interpreting the study findings, particularly unexpected findings.

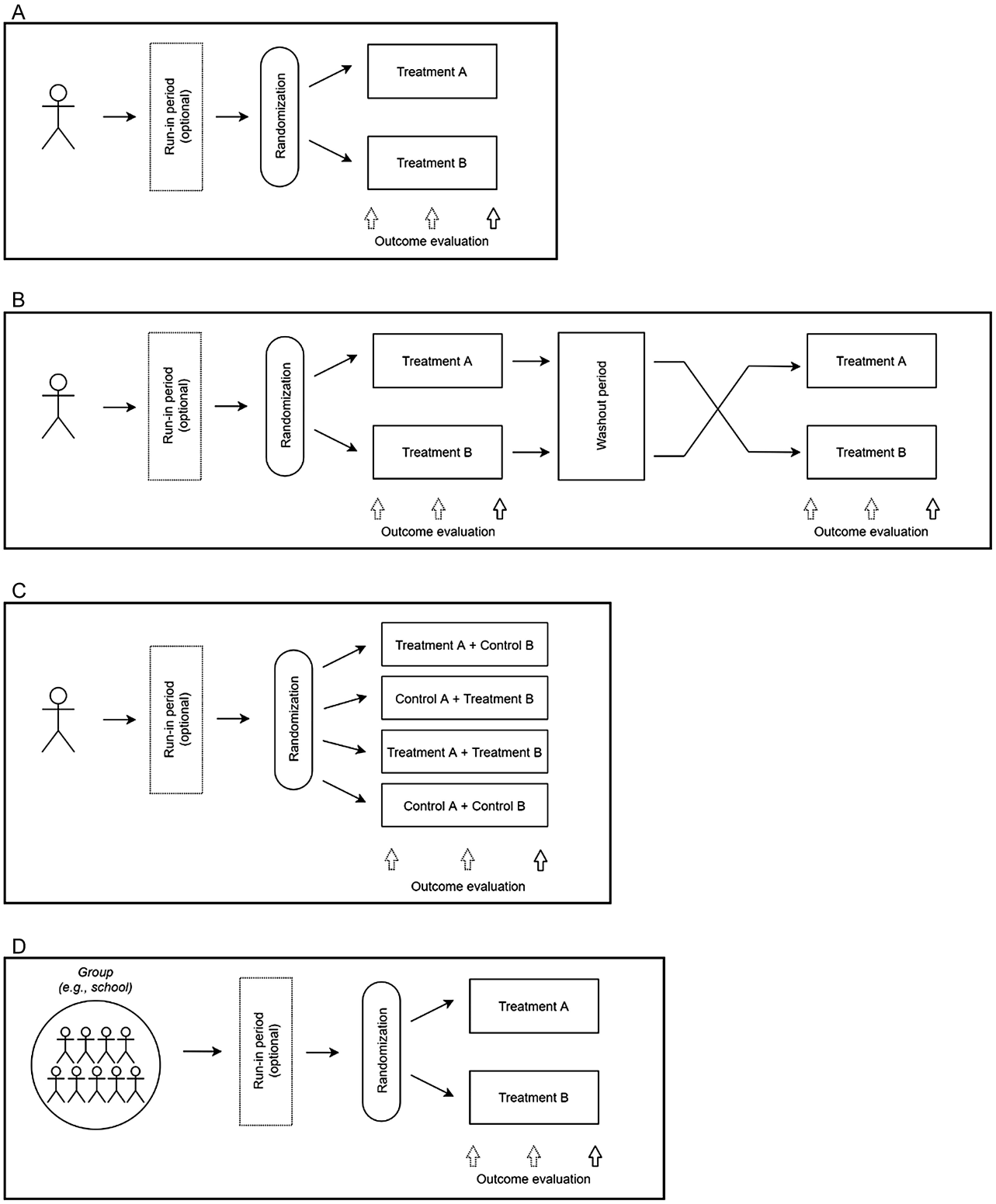

FIGURE 1 (A) Parallel design: participants may or may not participate in a run-in period, randomly assigned to treatment A or B for a 2-comparison design. (B) Crossover design: participants may or may not participate in a run-in period, randomly assigned to either treatment A followed by treatment B (after washout period) or treatment B followed by treatment A. (C) Factorial design: participants may or may not participate in a run-in period, participants randomly assigned to 1 of 4 groups; either treatment A or B and either control A or B. (D) Randomized cluster design: participants may or may not participate in a run-in period; all participants in 1 group (e.g., school) are assigned to treatment A or B.

Designing protocol foods/diets

Nutrient adequacy.

The Recommended Dietary Intakes, issued by the Food and Nutrition Board of the National Academy of Sciences, should guide decisions about nutrient adequacy. Several commercially available nutrient-composition databases are available for use to estimate the macronutrient profile and micronutrient content and allow for the management of 1 food/nutrient/dietary change, such as replacement, on the balance of the diet. It is important to note that nutrient data in commercially available nutrient composition databases are often derived from average composition values and, therefore, may not accurately reflect the specific study items. Therefore, the recommendation is to confirm the nutrient data by chemical analyses.

Acceptability/suitability of protocol food(s)/diet.

To ensure meeting targeted enrollment of study participants meeting the inclusion/exclusion criteria and high levels of adherence to the protocol (to the extent possible), study menus should reflect commonly consumed foods. Common allergens (e.g., peanuts) should be avoided. When possible, offering tastings of the main or unusual study foods during the screening process will acquaint potential participants to the items and minimize enrollment of individuals who are unlikely to be adherent. This step is particularly valuable for interventions of long duration. If the structure, matrix, or form of intervention foods and beverages is part of the research question, the team must consider how to ensure that the diet intervention maintains that form throughout the trial.

Designing protocol foods/menus.

When designing protocol foods/menus, the number of days in the food/menu rotation is an important consideration. Short cycles (e.g., 3 d) minimizes food and nutrient variability and weight fluctuations but can result in diet fatigue, leading to diminished adherence and increasing the risk of dropouts. A longer cycle (e.g., 1 wk) increases food preparation time, personnel training, procurement, and storage/inventory demands (refrigeration/freezer/dry space), and can potentially waste food items that have a limited shelf life. A 6-d rather than 7-d menu cycle avoids the repetition of the same foods/meals on the same day every week.

The intervention protocol will define the specifications of the food/diet. Targeted nutrient goals are initially determined by calculations using data from a nutrient database. The content should be verified by chemical analysis.

In controlled-feeding studies, menus are typically constructed for a range of energy increments (e.g., 1800, 2000, 2500 kcal/d to 4000 kcal/d) in 100-kcal/d to 300-kcal/d increments. Additionally, unit foods are an efficient way to adjust energy intake to address unintended body-weight gain or loss. Unit foods are designed to have a macro- and micronutrient composition consistent with that of the intervention diet, generally created in easily dispensable portions to (e.g., 100 calories per unit). As such, if a participant's energy needs to maintain body weight falls between 1 of 2 of the predesigned menu options (e.g., 2200 kcal), 2 unit foods can be added to the 2000 menu plan. Some examples include trail mix, baked muffins or bars.

Estimating energy requirements.

A variety of methods are available to estimate energy requirements, including the Harris-Benedict (32) and Mifflin-St Jeor equations (33). Regardless of the method used, regular monitoring of body weight and subjective assessment of hunger and fullness are important to ensure weight stability or the intended weight change, and adherence to the study protocol (31). In general, a 250-kcal deficit/d will result in half a pound (0.25 kg) of weight loss per week.

Quality control.

It is critical to ensure accurate delivery of the intervention throughout the study period. Approaches to maximize accuracy and minimize variability include procuring single batches of key foods/supplements, avoiding the use of highly perishable or seasonable items, and monitoring stability at the designated storage conditions. When it is impossible or impractical to procure a single batch, lot numbers and distribution dates should be tracked in the study database for use during data analysis.

Recruitment and screening

Rationale/justification for the study population.

The rationale for a targeted study population should be guided by the study's specific aims. If the target population is defined by multiple criteria [e.g., male, age 50–60 y, prehypertensive, BMI (kg/m^2) 20–25, not taking blood glucose–lowering medications), it may be challenging to recruit the target sample size. Another example of multiple, narrowly targeted criteria is the selection of individuals with a specific genetic profile. Multisite studies or other strategies should be considered. If the specific aims of the protocol requires a certain race or ethnicity, that information is often based on self-report (34, 35). Given the interest in understanding health disparities, especially with nutrition-related health outcomes, it is important to appreciate the complexities of self-reported information about race or ethnicity and allow for multiple responses. It should likewise be recognized that recruitment of a cohort with a composition reflective of the target population increases the generalizability of the final data. Investigators should be mindful of regulatory conditions specific for recruiting vulnerable populations (e.g., minors, employees, prisoners, wards of the state, pregnant women). As with all potential participants, requirements for obtaining informed consent (or assent) with these specific populations must be consistent with institutional review board (IRB) guidelines.

Participant eligibility criteria.

Inclusion and exclusion criteria of study participants are determined by the study-specific aims and enumerated in the

IRB-approved study protocol. Examples of general eligibility criteria include demographic characteristics, anthropometric measures, biochemical or clinical measures, and lifestyle behaviors (**Table 5**) (36). More-specific eligibility criteria may include relatively narrow ranges for clinical measures, such as blood pressure, blood chemistries, or body weight. Although in many cases the general population is of ultimate interest, due to the high level of variability independent of the intervention contributed by participant characteristics, it is usually necessary to limit recruitment to participants who share similar features and are within certain geographic areas for practical reasons. With regard to participant characteristics, the internal validity of a study is the degree to which a study is free from bias or systematic error (37). A study has internal validity if inferences from the study population reflect the inferences that would be observed in the entire population with similar characteristics. Importantly, a study's internal validity is a prerequisite for generalizability (also referred to as external validity). The importance of a thoughtfully defined study population cannot be overstated (38). Critical elements of participant recruiting, consenting, and study participation are defined in the CONSORT flow diagram (**Supplemental Figure 1**). Recruitment data must be reported on a yearly basis to the IRB, study sponsors (NIH enrollment statistics), and at the time of publication by the International Committee of Medical Journal Editors and many peer-reviewed journals.

Participant consent.

Consistent with IRB policy, an individual is deemed eligible to participate in a human nutrition RCT on the basis of the inclusion and exclusion criteria and his/her voluntary decision. Potential participants must be given adequate opportunity to review all study consent forms and ask questions. No study-related activity begins until the consent form is signed.

Participant recruitment.

Recruiting study participants is often one of the most challenging aspects of conducting a human nutrition RCT. All material used for participant recruitment must be approved by the IRB prior to use. Well-defined, planned, and IRB-approved screening strategies with adequate resources for implementation are critical for successful completion of a human nutrition RCT (39). Factors that impact recruitment rates include available funds for advertising, adequate staff to respond to study queries and to conduct screening visits, duration of study protocol, number and type of biological samples collected, number and length of study visits, nature of the intervention relative to usual diet and lifestyle behaviors, and eligibility criteria restrictiveness. The acceptability of the nutrition intervention, particularly in terms of diet composition, can impact both recruitment success and adherence to the study protocol. Adherence or the lack thereof to complex screening protocols frequently serves as a bellwether for subsequent adherence to study protocols.

Participant screening and enrollment.

All screening activities, including recruitment material, require an initial IRB-approved informed consent or Health Insurance Portability and Accountability Act waiver. The ideal screening protocol is efficient, cost-effective, and confers low burden to potential participants. In many cases, it is cost-efficient to use a multistage screening process—prescreen and full screen. This "funnel" approach starts with broad criteria so as to exclude ineligible individuals early in the process and, hence, minimize ineligible participant burden and use of resources.

Prescreening. The scope of a prescreening protocol should be relatively brief and limited in nature (minimum criteria necessary for determining potential eligibility). Initial approaches can include a review of medical records and/or participant databases. These activities, along with outreach recruitment strategies, such as advertisements, will identify a pool from which potential participants can be contacted. For the potential participant, the goal of the prescreening contact is to provide a brief study overview to gauge potential general interest and then eligibility. Variables that may be used for prescreening and can be collected remotely include self-reported body weight and height and food allergies/intolerances/dislikes. A brief prescreening visit can permit measurement of eligibility criteria, such as blood pressure or fasting blood glucose concentrations.

Full screening. If a potential participant qualifies on the basis of the prescreening criteria (or a prescreening protocol is deemed unnecessary), he/she can then be invited to attend a full-screening session, consistent with an IRB-approved study protocol. At that time, data required to determine eligibility are collected or confirmed (e.g., anthropometrics, demographics, medical history, blood pressure, blood measures). An important consideration for human nutrition RCTs is to obtain information on dietary habits, use of dietary supplements, and food aversions/allergies. The rigors of study participation should be discussed and emphasized during the screening visit, particularly, if appropriate, the burdens associated with controlled-feeding trials. Study-specific issues to discuss may include a participant's willingness to follow the study protocol, impact of the intervention on habitual social interactions, travel plans (business/vacation), special occasions (holidays/graduations), plans to relocate, food preparation responsibilities for household members, availability of secure storage space for study foods/beverages consistent with the specific conditions (e.g., heat, light), travel logistics for study visits, habitual physical activity patterns, and regular social demands. Allowing participants to sample some of the study foods or view a menu/list of foods can help identify potential adherence challenges. Inadequate communication of the study requirements and expectations is a disservice to the potential participant and increases the likelihood of study dropouts or nonadherence. The screening personnel should not minimize the inconveniences that are imposed by participation in a human nutrition RCT. To

TABLE 5 Potential recruitment criteria for human nutrition RCTs[1]

Criteria	Variables	Example of inclusion criteria	Examples of exclusion criteria
Demographic	Age	Age: 50–80 y	Values outside range
	Sex	Sex: women only	—
	Race/ethnicity	>30%: Hispanic or Latino	—
Anthropometric	Height, weight (BMI)	BMI: 25–35 kg/m^2	Values outside range
	Weight	BMI: 25–35 kg/m^2	Weight gain or loss >2 kg within prior 6 months
	Waist circumference	Waist circumference >35 inches (89 cm) in women and >40 inches (101.6 cm) in men	Values outside range
Biochemical and clinical	Blood biomarker values	Fasting glucose 100–200 mg/dL	Values outside range
	Medical history	—	Self-reported CVD (history of MI, stroke, heart failure, coronary artery bypass graft, stenosis >50%, angina, and PAD)
	Medication use	—	Hypercholesterolemia therapy, chemotherapy, hypoglycemia therapy
	Menopausal status	Complete natural cessation of menses ≥12 mo or a bilateral oophorectomy	Premenopausal
	Health status	—	Renal or kidney disease (glomerular filtration rate <60 mL/(min · 1.73m^2) (caveat: normal GFR varies according to age, sex, and body size, and declines with age[2]); hypothyroidism or hyperthyroidism (TSH <0.4 or >4.5); type I and II diabetes
Lifestyle	Habitual dietary intake	—	Food allergies, aversions, cultural or religious dictates that restrict intake selected food/beverage items
	Habitual level of physical activity	—	Anticipated dramatic change in physical activity pattern (e.g., initiate marathon training)
	Tobacco use	—	Includes chewing tobacco, "snuff," nicotine gum and patches
	Habitual alcohol consumption	—	>1 drink/d in females and >2 drinks/d in males; if protocol specific, refusal to abstain during intervention period (1 drink = 14 g pure alcohol)
	Habitual supplement use	—	Supplement use within past 3 or 6 mo and, if protocol specific, refusal to discontinue during intervention period
Study-specific issues	Willingness to maintain body weight during intervention period	—	Plans to reduce weight (unless part of protocol)
	Commitment to intervention period	—	Plans to relocate, extended travel
	Plans to become pregnant or schedule elective surgery	—	Pregnancy or plans to become pregnant; plans to schedule elective surgery
	Interaction with study personnel	—	Investigator discretion
Other	Logistical determinants	—	Inadequate transportation options for study visits or storage (refrigerator, freezer) for intervention material, housing insecurity
	Simultaneous participation in another research study	—	Major food preparation responsibilities for household
	Social Security number	—	Stipend payment prohibited if no Social Security number

[1] CVD, cardiovascular disease; GFR, glomerular filtration rate; MI, myocardial infarction; PAD, peripheral artery disease; RCT, randomized controlled trial; TSH, thyroid stimulating hormone.
[2] GFR Calculator/National Kidney Foundation (40).

the extent possible, it is advisable to minimize the duration between screening and enrollment of eligible participants so that they do not lose interest or need to be rescreened due to measures that are time sensitive (e.g., pregnancy tests).

Challenges to participant recruitment.
Defining eligibility criteria often involves a trade-off between scientific and practical goals. Recruitment might be limited due to restrictive inclusion criteria, such as a cutoff value for a clinical characteristic (meeting 5 metabolic syndrome criteria) or use of an excluded medication (e.g., blood pressure, blood lipids). In such instances, modifying the eligibility criteria (meeting 3 of 5 metabolic syndrome criteria) or medication use (specifying that the participant be normotensive without or with medication) can be considered, dependent on the study-specific aims. All changes to the eligibility criteria must be approved by the IRB prior to implementation. Regular meetings of the study team to review recruitment goals and confirm eligibility of each participant help identify and address potential recruitment issues in a timely manner. Including nonspecific eligibility criteria in the protocol, such as investigator "discretion," permits exclusion of participants who are unlikely, in the opinion of the study team, to complete the study or be compliant.

Participant burden.
Heavy participant burden can negatively impact participant adherence and retention. This is a particular risk for studies that involve a large team of investigators, each with their own research agendas. The final decision on outcome measures should be based on a balance between study-specific aims and participant burden, particularly in terms of perceptions that the assessments are safe and acceptable or overly intrusive, time consuming, and tedious. Pilot testing of complex study designs is advisable. Studies should be designed to optimize benefit and minimize burden to participants.

Establishing a study stipend.
Remuneration to study participants is commonplace and may be institution specific. Investigators should determine an appropriate level of compensation that upholds respect for the participants and ensures recruitment and retention of study participants, while avoiding coercion or the perception of coercion (41). Stipend guidelines must be approved and at times are established by IRBs.

Ethical considerations.
Designing human nutrition RCTs can raise unique ethical questions: for example, withholding a nutrient that is thought to be low or deficient in study participants or identifying an ethical comparator or control substance (28). At each step in the study design process these and similar issues should be critically identified, discussed, addressed, and when necessary, input solicited from the IRB.

The study protocol should include plans to inform participants about abnormal clinical, imaging, or other results that are identified during the screening or implementation phases of the RCT. When possible, abnormal results should be provided to participants in person or via phone. It is likewise strongly advised to encourage participants to follow up with their primary care provider about abnormal results.

Initiation of Human Nutrition RCTs

Randomization

Randomization is a critical component of nutrition intervention studies, intended to minimize conscious and unconscious biases, and increase the probability that individuals in each treatment sequence or group have similar baseline characteristics. Randomization and blocking are intended to distribute variables (known and unknown confounders) that might affect study outcomes equally across groups, such as age, sex, and race. Randomization can be stratified to ensure an equal distribution of individuals in predefined subgroups, particularly when a baseline variable is expected to influence study outcomes. Simple randomization of participants can be used for all study designs where the individual is the study unit and individuals are considered independent (e.g., parallel, crossover, and factorial designs). If there are any dependencies in the data caused by shared environmental, behavioral, or genetic factors, randomization must account for dependency between/among participants. For example, multiple members of the household may screen and qualify for human nutrition RCT participation. In such cases, in the absence of specific exclusions on the basis of household residency, it is advisable to conduct randomization by household and appropriately account for the intrahousehold correlation in the statistical plan. Clarity on this issue is paramount as has recently been demonstrated (42, 43).

The allocation sequence can be generated using an online statistical computing Web program (44) or by programming the randomization design in a statistical software package [R (R Foundation for Statistical Computing), SAS (SAS Institute), or Python] (45). Two methods to minimize potential for bias and confounding are block randomization and stratified randomization. Block randomization allocates participants within small aggregates (blocks), such that an equal number is assigned to each treatment. Stratified randomization facilitates achieving balance in the allocation of participants to treatment groups and minimizes unbalanced groups (e.g., a priori plan for an equal distribution of women and men, age ranges, BMI ranges) (6).

Intervention blinding

Intervention blinding is a critical aspect of human nutrition RCTs intended to minimize unintentional bias to the extent possible. For a single-blind study design, study team members, but not participants, are naive to the intervention allocation. The necessity of this approach is determined by the nature of the intervention. Unlike drug

trials, human nutrition RCTs frequently test foods that are identifiable by taste, appearance, texture, and/or smell. Thus, the formulation of a "matched placebo" is not feasible. Examples of scenarios where a single-blind intervention may be necessary include comparisons of animal versus plant proteins, solid fats versus liquid fats, and simple sugars versus complex carbohydrates. To ensure the study team is unaware of the intervention allocation (with the exception of the staff who prepare and dispense the food), no identifiers should be incorporated into the labeling of diet components and biological samples (46).

For a double-blind study design, both participants and study team members are naive to the intervention allocation. A double-blind study is possible when the interventions can be presented in a way that appears similar or identical to the comparator and the variable component of the diet can be incorporated (hidden) into food mixtures. Examples of scenarios where a double-blind intervention may be possible include comparisons of types of liquid vegetable oils incorporated into muffins or types of vegetable proteins incorporated into tomato sauces. To ensure intervention blinding for the study team, with the exception of those who prepare and dispense the food, no identifiers should be incorporated into the labeling of diet components and biological samples. For a triple-blind study, personnel who assess the outcomes (analysts and statisticians) also remain unaware of treatments until the analyses are complete (47).

Run-in period

A run-in period may be incorporated into the study design. If it is part of the protocol it occurs following screening but before baseline measures are collected and is generally shorter than the intervention period(s). During this period, all participants are given the same diet. A run-in period allows investigators to exclude participants who cannot comply with the diet prior to randomization. However, this approach may introduce some selection bias (48). A run-in period also serves to reduce variance inherent in the outcome measurements (reduces the SD) and attenuates regression to the mean (reversion to the mean and reversion to mediocrity), the convergence to the population mean with repeated measurements. A run-in period is commonly used in trials of placebo-controlled pharmaceutical studies (49). Of note, a run-in period for a human nutrition RCT may improve participants' health status, because the diet is healthier than the participant's habitual diet, even when that diet is designed to reflect the composition of an average US diet (50). Therefore, a run-in period may attenuate improvements in outcome measurements with the test diets. It also may introduce order effects in crossover studies because the run-in period is always completed prior to the first diet period.

Adherence and retention

Adherence and compliance are terms used to describe a participant's ability to follow an investigator's instructions about the use of medications, regimens, diets, and/or systems. The word "adherence" is the preferred terminology over "compliance," because compliance may suggest passively following instructions (51). **Table 6** provides examples of real-time adherence assessment, adherence classification, and adherence biomarkers. Prior to starting the study, investigators should develop protocols to measure, define, and manage adherence. For example, adherence could be measured using pill counts, assessed by requiring and monitoring return of unused supplements, monitoring 24-h excretion of a dietary component (e.g., sodium, potassium), or spiking a sample with a biomarker and monitoring 24-h excretion (e.g., para-aminobenzoic acid, riboflavin). See **Supplemental Table 3** for examples of issues related to nonadherence and possible solutions.

A study protocol should define how nonadherence would be handled during data analysis. In efficacy trials, nonadherent participants may be withdrawn from the study. However, in effectiveness trials, nonadherent participants are frequently retained to evaluate the feasibility of the intervention under real-world settings. In effectiveness trials, investigators may conduct a per-protocol analysis. To enhance adherence and retention, it is helpful to identify potential pitfalls that could compromise adherence. Identifying potential pitfalls can inform the development of strategies to overcome barriers to adherence and minimize protocol deviations. Maintaining participant enthusiasm and investment is critical. Effective communication promotes a strong sense of participant allegiance and enforces the importance of adherence to study protocol.

Sample size considerations

Calculation of sample size.

Calculating sample size is a critical aspect of study design and is dependent on variance and effect magnitude of the outcome(s) of interest (52). Underestimating sample size will result in inadequate statistical power, while overestimating sample size is costly and wastes resources. Statistical power is the probability of detecting an effect when the effect truly exists. The type of calculation itself depends on the planned statistical test for the primary outcomes (53). For example, a 2-group parallel design to study change in blood pressure over 4 wk can be tested with a 2-sample Student's t test, and the sample size can be estimated using formulas based on an unpaired t test. For a 3-treatment crossover design that will be analyzed using repeated-measures ANOVA, a simulation based on repeated-measures ANOVA should be used to estimate sample size. The key inputs for sample size calculations include the clinically relevant detectable change in the study outcome, variation of the outcome, type I error, and statistical power. Cluster-randomized designs additionally require an input for intra-cluster correlation.

Sample sizes can be estimated using statistical software or simulation. Programs such as SAS software and R have routines for standard models, while specialized sample size software, such as G∗Power (54) and Power Analysis and

TABLE 6 Strategies to monitor adherence in real time and for per-protocol analyses[1]

Assessment method	Contribution to adherence detection	Level used to classify adherence vs nonadherence
Real-time adherence assessment		
Food consumption in the presence of study personnel	Participants consume all or some (e.g., feeding studies, 1 main meal may be consumed onsite) of the total study food(s) in the presence of the study personnel.	Calculate percent study food consumed. Criteria for nonadherence should be established prior to study initiation.
Monitor body weight	If protocol is designed to be energy-balanced, weight should be relatively stable; if designed to be energy-deficient, weight should decline.	Assuming provided energy is accurately matched to energy expenditure (resting energy expenditure and physical activity), body-weight change anomalies indicate nonadherence. Criteria for nonadherence should be established prior to study initiation.
Daily/weekly monitoring forms	Participant reports deviations in intake [e.g., study food(s)/beverage(s) not consumed or non–study food(s)/beverage(s) consumed].	Days when deviation from the protocol is reported classified as nonadherent. Criteria for nonadherence should be established prior to study initiation.
Phone call/text/e-mail	Spot checks regarding questions/clarifications about adherence.	Criteria for nonadherence should be established prior to study initiation.
Adherence classification for per-protocol analyses		
Daily monitoring or weekly monitoring forms	Deviations from protocol noted [e.g., study food(s)/beverage(s) not consumed or non–study food(s)/beverage(s) consumed].	Days where any deviation from the protocol is reported are classified as nonadherent. Prespecify criteria for triggering study withdrawal due to nonadherence (e.g., ≤90% of study days).
Dietary intake assessment	Data for the whole diet or consumption of specific foods captured.	Prespecify criteria for triggering study withdrawal due to nonadherence.
Urinary biomarker excretion		
Dietary components that have near complete excretion	24-h urinary sodium, potassium	Criteria for nonadherence should be established prior to study initiation.
Compounds added to study food(s) that have known percent excretion	24-h urinary PABA, riboflavin	Criteria for nonadherence should be established prior to study initiation.

[1] PABA, para-aminobenzoic acid.

Sample Size Software (PASS) (55), have simulation routines for complex models (56). If treatment effects within a subgroup is the primary objective of the study, then sample size calculations should be applied to each subgroup. If multiple study sites are involved, this variable should be accounted for in the sample size calculation using an intraclass correlation coefficient estimate to calculate an effective sample size. Final sample size numbers should account for expected attrition and nonadherence.

Power calculations can be provided for all secondary outcome measures that have a priori hypotheses. These secondary outcomes are often not included in the experiment-wise type I error allocated for the primary outcome(s). If there are a large number of secondary outcomes, the type I error across all secondary outcomes can be adjusted for multiple testing as a whole, or in groups of related outcomes. For example, a group of 6 proinflammatory cytokines may correspond to one of the secondary hypotheses. If power calculations are provided for these outcomes, a Bonferroni multiple-comparison adjustment can be applied to the type I error in the power calculation.

Assessment of detectable change.

Detectable change is the smallest change in the outcome that a study is statistically powered to detect. Ideally, the detectable change represents a clinically meaningful change in the primary outcome(s). Investigators can relate a meaningful change to improved quality of life, lowered event rates, lowered morbidity or mortality rates, or lower health costs. It is important to cite prior work linking detectable change to a clinically meaningful hard outcome(s). Likewise, it is critical to demonstrate the feasibility of detecting the predicted change using the planned study design and intervention (57).

Determining variation in the outcome.

Variation in the outcome depends on the statistical analysis plan. For example, if the planned analysis is an unpaired *t* test, then a between-participant estimate of variance is required. If the planned analysis is a paired *t* test, then a within-participant estimate of variance is required. Variance estimates can be obtained from published studies or unpublished preliminary data. The best estimate of variance will reflect the target population of interest, thus will often come from a previous study with similar inclusion criteria and a similar intervention period. When testing differences in means, group SDs of the outcome are needed. Publications will often report a table with pre- and postintervention means and SEs. Within-group SEs can be converted to SD by multiplying by the square root of the group sample size. Nonnormally distributed preintervention and postintervention outcomes may have SDs related to the mean [e.g., triglyceride and

Lp(a) concentrations] and will need to be assessed using nonparametric sample size methods or a sample size calculated using a different scale (under an appropriate variable transformation). For crossover studies, if previously reported variances are not available, within-participant variation can be estimated from between-participant variation along with an assumption on the within-participant correlation. If there are no preliminary data available, an SD can be derived based on the range of plausible values (convert to SD by dividing the range by 4) or a standardized effect size. Alternatively, Cohen's *d* (calculated by taking the difference in the group means and dividing by the pooled SD) can be used (58).

Type I error.

Type I error or family-wise type I error [or family-wise error rate (FWER)] is the probability of detecting an effect when the null hypothesis is true (rejecting the true null hypothesis). If multiple hypotheses are tested, the FWER reflects the probability of having $\geq$1 false-positive results. The risk of a type I error increases with multiple hypothesis testing (e.g., high-dimensional 'omics data). Multiple hypothesis testing scenarios need to be identified a priori and accounted for in the type I error specification. Common multiple testing adjustment procedures include Bonferroni adjustment for small numbers of tests and false discovery rate adjustment for large numbers of tests. For example, if there are 2 co-primary outcomes, the experiment-wise type I error can be limited to 0.05 by using $\alpha = 0.025$ in each of the individual calculations. In trials with >2 treatment groups, where pairwise comparisons are of primary interest, the number of comparisons can be taken into account using a multiple testing adjustment method (57).

Type II error.

Type II error is the probability of not rejecting the null hypothesis when it is false (false negative). The probability of a type II error represents 1 minus the power of the test (B error). Control for type II error is only needed when a study has multiple endpoints that all need to be found significant in order to determine a treatment effect. Such a scenario is most commonly found in Food and Drug Administration efficacy reviews and is uncommon in nutrition studies.

Statistical power

By convention, power should be at least 80%, although the threshold of 90% is frequently used.

Analysis plan

Investigators must describe in detail the statistical analysis plan in the study protocol. The analysis plan should explain how study results can be linked to a priori stated outcomes, thereby defining the impact of the study. The analysis plan should account for the expected amount of time and personnel necessary to complete statistical analyses. Prior to the start of the study, it is critical to verify that the protocol includes collection of the precise data necessary to address the study specific aims. Analysis techniques are described in the third paper of the series (18).

Summary

Human nutrition RCTs have been, and will continue to be, critical components of formulating evidence-based dietary guidance. Frequently, efforts to formulate dietary guidance have been hampered by a limited number of studies of sufficient quality to contribute to the evidence base on which recommendations can be constructed. Addressed in this report are major issues that should be considered when planning and conducting human nutrition RCTs. Addressed are issues related to reporting guidelines; constructing research questions and specific aims; distinguishing among types of study designs; choosing biomarker and outcome measures; collecting and archiving study data and samples; designing intervention protocols; initiating participant recruitment and screening; considering participant eligibility, burden, and retention; designing and conducting the RCT; and estimating sample size relative to statistical power. All of these study components will affect the utility and generalizability of the study results.

Acknowledgments

The authors thank all the organizations that provided financial support, resource support, and leadership on this project including the Tufts Clinical and Translational Science Institute (CTSI), Indiana CTSI, and Penn State CTSI. The authors are also grateful to Tufts CTSI for providing overall leadership on this project, organizing all writing group meetings, providing project management support, and hosting the writing workshop that initiated this project. The authors' responsibilities were as follows—AHL, KP, KB, KEH, CAMA, DJB, JWL, HR, and NRM: drafted the manuscript and had responsibility for the final content; AHL: edited and revised the manuscript; and all authors: read and approved the final manuscript.

References

1. 2015 Dietary Guidelines Advisory Committee. Advisory report to the Secretary of Health and Human Services and the Secretary of Agriculture. Washington (DC): US Department of Agriculture, Agricultural Research Service; 2015.
2. EQUATOR Network [Internet]. [Cited June 15, 2020]. Available from: https://www.equator-network.org/reporting-guidelines/.
3. CONSORT Transparent Reporting of Trials. [Internet]. [Cited June 15, 2020]. Available from: http://www.consort-statement.org/checklists/view/32–consort-2010/66-title.
4. Hulley SB. Designing clinical research. 3rd ed. Philadelphia (PA): Lippincott Williams & Wilkins; 2007.
5. Chow S-C, Liu J-P. Design and analysis of clinical trials: concepts and methodologies. Germany (DEU): John Wiley & Sons; 2008.
6. Efird J. Blocked randomization with randomly selected block sizes. Int J Environ Res Public Health 2011;8(1):15–20.
7. O'Connor LE, Li J, Sayer RD, Hennessy JE, Campbell WW. Short-term effects of healthy eating pattern cycling on cardiovascular disease risk factors: pooled results from two randomized controlled trials. Nutrients 2018;10(11):1725.

8. Jones B, Kenward MG. Design and analysis of cross-over trials. 3rd ed. CRC Press, Taylor & Francis Group: Chapman and Hall/CRC; 2014[Internet]. Available from: https://www.routledge.com/Chapman--HallCRC-Texts-in-Statistical-Science/book-series/CHTEXSTASCI.
9. Ginsberg HN, Kris-Etherton P, Dennis B, Elmer PJ, Ershow A, Lefevre M, Pearson T, Roheim P, Ramakrishnan R, Reed R. Effects of reducing dietary saturated fatty acids on plasma lipids and lipoproteins in healthy subjects: the DELTA study, protocol 1. Arterioscler Thromb Vasc Biol 1998;18(3):441–9.
10. Berglund L, Lefevre M, Ginsberg HN, Kris-Etherton PM, Elmer PJ, Stewart PW, Ershow A, Pearson TA, Dennis BH, Roheim PS. Comparison of monounsaturated fat with carbohydrates as a replacement for saturated fat in subjects with a high metabolic risk profile: studies in the fasting and postprandial states. Am J Clin Nutr 2007;86(6):1611–20.
11. Appel LJ, Moore TJ, Obarzanek E, Vollmer WM, Svetkey LP, Sacks FM, Bray GA, Vogt TM, Cutler JA, Windhauser MM. A clinical trial of the effects of dietary patterns on blood pressure. N Engl J Med 1997;336(16):1117–24.
12. Appel LJ, Sacks FM, Carey VJ, Obarzanek E, Swain JF, Miller ER, Conlin PR, Erlinger TP, Rosner BA, Laranjo NM. Effects of protein, monounsaturated fat, and carbohydrate intake on blood pressure and serum lipids: results of the OmniHeart randomized trial. JAMA 2005;294(19):2455–64.
13. Sacks FM, Carey VJ, Anderson CA, Miller ER, Copeland T, Charleston J, Harshfield BJ, Laranjo N, McCarron P, Swain J. Effects of high vs low glycemic index of dietary carbohydrate on cardiovascular disease risk factors and insulin sensitivity: the OmniCarb randomized clinical trial. JAMA 2014;312(23):2531–41.
14. Manson JE, Cook NR, Lee I-M, Christen W, Bassuk SS, Mora S, Gibson H, Gordon D, Copeland T, D'Agostino D. Vitamin D supplements and prevention of cancer and cardiovascular disease. N Engl J Med 2019;380(1):33–44.
15. Manson JE, Bassuk SS, Lee I-M, Cook NR, Albert MA, Gordon D, Zaharris E, MacFadyen JG, Danielson E, Lin J. The VITamin D and OmegA-3 TriaL (VITAL): rationale and design of a large randomized controlled trial of vitamin D and marine omega-3 fatty acid supplements for the primary prevention of cancer and cardiovascular disease. Contemp Clin Trials 2012;33(1):159–71.
16. Howard BV, Van Horn L, Hsia J, Manson JE, Stefanick ML, Wassertheil-Smoller S, Kuller LH, LaCroix AZ, Langer RD, Lasser NL. Low-fat dietary pattern and risk of cardiovascular disease: the Women's Health Initiative Randomized Controlled Dietary Modification Trial. JAMA 2006;295(6):655–66.
17. Appel LJ., Champagne CM, Harsha DW, Cooper LS, Obarzanek E, Elmer PJ, Stevens VJ, Vollmer WM, Lin PH, Svetkey LP. Effects of comprehensive lifestyle modification on blood pressure control: main results of the PREMIER clinical trial. JAMA 2003;289(16):2083–93.
18. Maki KC, Miller JW, McCabe GP, Raman G, Kris-Etherton PM. Perspective: Laboratory considerations and clinical data management for human nutrition randomized controlled trials: Guidance for ensuring quality and integrity. Adv Nutr 2021;12(1):46–58.
19. Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, McManus K, Champagne CM, Bishop LM, Laranjo N. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. N Engl J Med 2009;360(9):859–73.
20. Gardner CD, Trepanowski JF, Del Gobbo LC, Hauser ME, Rigdon J, Ioannidis JP, Desai M, King AC. Effect of low-fat vs low-carbohydrate diet on 12-month weight loss in overweight adults and the association with genotype pattern or insulin secretion: the DIETFITS randomized clinical trial. JAMA 2018;319(7):667–79.
21. Chew EY, Clemons TE, SanGiovanni JP, Danis R, Ferris FL, Elman M, Antoszyk A, Ruby A, Orth D, Bressler S. Lutein+ zeaxanthin and omega-3 fatty acids for age-related macular degeneration: the Age-Related Eye Disease Study 2 (AREDS2) randomized clinical trial. JAMA 2013;309(19):2005–15.
22. Neal B, Tian M, Li N, Elliott P, Yan LL, Labarthe DR, Huang L, Yin X, Hao Z, Stepien S. Rationale, design, and baseline characteristics of the Salt Substitute and Stroke Study (SSaSS)—a large-scale cluster randomized controlled trial. Am Heart J 2017;188:109–17.
23. Kaur J, Kaur M, Webster J, Kumar R. Protocol for a cluster randomised controlled trial on information technology-enabled nutrition intervention among urban adults in Chandigarh (India): SMART eating trial. Glob Health Action 2018;11(1):1419738.
24. Delaney T, Wyse R, Yoong SL, Sutherland R, Wiggers J, Ball K, Campbell K, Rissel C, Wolfenden L. Cluster randomised controlled trial of a consumer behaviour intervention to improve healthy food purchases from online canteens: study protocol. BMJ Open 2017;7(4):e014569.
25. Staudacher HM, Irving PM, Lomer MCE, Whelan K. The challenges of control groups, placebos and blinding in clinical trials of dietary interventions. Proc Nutr Soc 2017;76(3):203–12.
26. Mensink RP. Effects of saturated fatty acids on serum lipids and lipoproteins: a systematic review and regression analysis. Geneva (Switzerland): World Health Organization; 2016.
27. Li Y, Hruby A, Bernstein AM, Ley SH, Wang DD, Chiuve SE, Sampson L, Rexrode KM, Rimm EB, Willett WC. Saturated fats compared with unsaturated fats and sources of carbohydrates in relation to risk of coronary heart disease: a prospective cohort study. J Am Coll Cardiol 2015;66(14):1538–48.
28. Weaver CM, Miller JW. Challenges in conducting clinical nutrition research. Nutr Rev 2017;75(7):491–9.
29. Bailey R, Weaver C, Murphy S. Using the Dietary Reference Intakes to assess intakes.In: Research: successful approaches. Chicago: Academy of Nutrition and Dietetics; 2018 , J Am Diet Assoc 2006;106(10):1550–3.
30. Kranz S, Hill AM, Fleming JA, Hartman TJ, West SG, Kris-Etherton PM. Nutrient displacement associated with walnut supplementation in men. J Hum Nutr Diet 2014;27(s2):247–54.
31. Steiber AL, Hand RK, Papoutsakis C. Guidelines for developing and implementing clinical nutrition studies.In: Van Horn L, Beto J, editors. Research: successful approaches in nutrition and dietetics. Chicago: Academy of Nutrition and Dietetics; 2019.
32. Frankenfield DC, Muth ER, Rowe WA. The Harris-Benedict studies of human basal metabolism: history and limitations. J Am Diet Assoc 1998;98(4):439–45.
33. Mifflin MD, St Jeor ST, Hill LA, Scott BJ, Daugherty SA, Koh YO. A new predictive equation for resting energy expenditure in healthy individuals. Am J Clin Nutr 1990;51(2):241–7.
34. Mersha TB, Abebe T. Self-reported race/ethnicity in the age of genomic research: its potential impact on understanding health disparities. Hum Genomics 2015;9(1):1.
35. Amorrortu RP, Arevalo M, Vernon SW, Mainous AG, Diaz V, McKee MD, Ford ME, Tilley BC. Recruitment of racial and ethnic minorities to clinical trials conducted within specialty clinics: an intervention mapping approach. Trials 2018;19(1):115.
36. Gordis L. Epidemiology. 5th ed. Elsevier: Saunders Press; 2013.
37. Porta M. A dictionary of epidemiology. 5th ed. Elsevier: Oxford University Press; 2008.
38. Hulley SB, Cummings SR, Browner WS, Grady D, Newman TB. Designing clinical research. 4th ed. Elsevier: Wolters Kluwer/Lippincott Williams & Wilkins; 2013.
39. Kris-Etherton PM, Mustad V, Lichenstein A. Recruitment and screening of study participants. In: Dennis B,editor. Well-controlled diet studies in humans: a practical guide to design and management. Chicago: American Dietetic Association; 1999. p. 76–96.
40. National Kidney Foundation. GFR calculator [Internet]. Available from: https://www.kidney.org/professionals/KDOQI/gfr_calculator.
41. Czarny M, Kass NE, Flexner C, Carson KA, Myers R, Fuchs E. Payment to healthy volunteers in clinical research: the research subject's perspective. Clin Pharmacol Ther 2010;87(3):286–93.
42. Estruch R, Ros E, Salas-Salvadó J, Covas M-I, Corella D, Arós F, Gómez-Gracia E, Ruiz-Gutiérrez V, Fiol M, Lapetra J. Primary prevention of cardiovascular disease with a Mediterranean diet. N Engl J Med 2013;368(14):1279–90.
43. Estruch R, Ros E, Salas-Salvado J, Covas MI, Corella D, Aros F, Gomez-Gracia E, Ruiz-Gutierrez V, Fiol M, Lapetra J, et al. Primary prevention of cardiovascular disease with a mediterranean diet supplemented

with extra-virgin olive oil or nuts. N Engl J Med 2018;378(25): e34.
44. Dallal GE. Homepage [Internet]. Available from: http://www.randomization.com.
45. Suresh K. An overview of randomization techniques: an unbiased assessment of outcome in clinical research. J Hum Reprod Sci 2011;4(1):8–11
46. Friedman LM, Furberg C, DeMets DL. Fundamentals of clinical trials. New York: Springer; 1998.
47. Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. Ann Intern Med 2002;136(3):254–9.
48. Boushey C. Van Horn L, Beto J. Building the research foundation: the research question and study design. In:, editors. Research: successful approaches in nutrition and dietetics. Chicago: Academy of Nutrition and Dietetics; 2019.
49. Packer M. Why has a run-in period been a design element in most landmark clinical trials? Analysis of the critical role of run-in periods in drug development. J Card Fail 2017;23(9):697–9.
50. Tindall AM, Petersen KS, Kulas-Ray ACS, Richter CK, Proctor DN, Kris-Etherton PM. Replacing saturated fat with walnuts or vegetable oils improves central blood pressure and serum lipids in adults at risk for cardiovascular disease: a randomized controlled-feeding trial. J Am Heart Assoc 2019;8(9):e011512.
51. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005;353(5):487–97.
52. Lucey A, Heneghan C, Kiely ME. Guidance for the design and implementation of human dietary intervention studies for health claim submissions. Nutr Bull 2016;41(4):378–94.
53. Chow S-C, Shao J, Wang H, Lokhnygina Y. Sample size calculations in clinical research. Chapman and Hall/CRC; 2017.
54. Faul F, Erdfelder E, Lang A-G, Buchner A. G∗ Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39(2): 175–91.
55. NCSS Statistical Software. 2019 PASS Power Analysis and Sample Size Software. Kaysville (UT): NCSS statistical software; 2019 [Internet]. Available from: https://www.ncss.com/software/pass/.
56. R Core Team. R: a language and environment for statistical computing [computer software]. Vienna (Austria): R Foundation for Statistical Computing; 2017.
57. Brooks G, Johanson G. Sample size considerations for multiple comparison procedures in ANOVA. J Mod Appl Stat Methods 2011;10(1):97–109.
58. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol 2005;5(1):13.