3
Study Design: Intervention Studies

Jayne V Woodside,[1] Robert W Welch,[2] Chris C Patterson[1] and Michelle C McKinley[1]

[1] *Centre for Public Health, Queen's University Belfast*
[2] *Northern Ireland Centre for Food and Health, School of Biomedical Sciences, University of Ulster*

Key messages

- Substantial evidence links nutrition to the improvement of physiological function and/or the reduction of the risk of major chronic diseases.
- Intervention studies fulfil an important role in establishing the link between nutrition and improvements in health as, if well designed, they may allow the testing of causality.
- Nutrition intervention studies can range in type from a double-blind, randomised, placebo-controlled, nutrient supplementation study through to a community-based lifestyle intervention, or a population-based fortification project.
- Nutrition intervention studies can range in duration, from short-term studies assessing acute postprandial effects of specific dietary modifications through to long-term interventions running over many months or years that examine change in risk markers or incidence of disease.
- The study design process should include careful consideration of the hypothesis, duration, intervention, amount and mode of delivery, control and blinding, primary and secondary outcome measures (including assessment of background diet), statistical power, eligibility criteria, data-collection methodology and ways of measuring and encouraging compliance.
- Advice from a statistician during both study design and statistical analysis is recommended.
- Local ethical approval and research governance procedures must be followed, and intervention studies registered on a publicly accessible database before recruitment commences. Any potential conflicts of interest, for example when funding has come from the food industry, should be declared.

3.1 Introduction

There is substantial evidence linking dietary factors to the primary and secondary prevention of major chronic diseases such as heart disease, diabetes and certain cancers, as well as the improvement of physiological function and the maintenance of adequate nutritional status.

Although observational studies (see Chapter 2) can demonstrate an association between a particular nutrient, food or diet and a functional or disease-related endpoint, causality cannot be demonstrated using such study designs. To demonstrate cause and effect requires an intervention study in which consumption of a nutrient, food or diet is altered in a controlled way and the effect on selected outcomes is measured. Intervention studies are higher up the hierarchy of scientific evidence than observational studies, although a combination of different study designs is usually utilised to develop a comprehensive evidence base for a link between consumption of a particular food or nutrient and a health-related outcome. Observational studies often generate hypotheses, which can be tested more rigorously in an intervention study.

This chapter will examine the different types of intervention study and then outline some of the key factors to consider when planning such studies. It includes intervention study design when the focus of interest is a particular nutrient, whole food, food group or whole diet, and also discusses nutrient supplementation studies.

3.2 Intervention study types

Intervention studies should be hypothesis driven and have a strong evidence basis. Intervention study designs can range from a short-term study, where the immediate

Nutrition Research Methodologies, First Edition. Edited by Julie A Lovegrove, Leanne Hodson, Sangita Sharma and Susan A Lanham-New.

Companion Website: www.wiley.com/go/nutritionsociety

effect of the intervention (consumed once) is measured over minutes to hours (for example, postprandial, or post-meal, studies), through to long-term studies that evaluate the effects of the intervention over a period of weeks, months or years. The study setting can also vary, from those in free-living populations through to studies conducted entirely in purpose-built research facilities such as metabolic suites, or within clinical facilities such as metabolic wards. The main study designs are outlined in this section.

Pilot studies

Different definitions exist for a pilot study (sometimes the terms 'feasibility study' or 'exploratory study' are also used interchangeably), but they are generally regarded as studies that are implemented on a small scale in order to test whether all the study processes operate as anticipated before undertaking a full-scale trial. There are many different reasons for performing pilot studies: for example, they are often undertaken in order to assess how realistic it would be to conduct a full-scale trial, to test recruitment procedures in a defined population, to develop and test research instruments, to develop and test a novel intervention (e.g. to evaluate food matrix issues or to ascertain dose or amount to be consumed), to identify logistical challenges in implementing a full-scale trial and to convince funding bodies that such a trial is worth funding. These studies can also provide data on the distribution/variability and timescale of outcome responses, which can be used for power calculations in subsequent definitive studies. Pilot studies vary in design and may test all, or only some, aspects of a full-scale study. They may be single-arm (before and after) studies with no control group, and these can be a cost- and time-effective way of assessing potential effects, but only as a forerunner to controlled studies. Pilot studies add to the totality of evidence, but on their own cannot determine the effect of intervention.

In general, data from pilot studies should be reported in descriptive terms and caution should be exercised when interpreting any statistical tests of significance, which will typically lack power. Publishing pilot data provides important insights and information that can be used by other researchers and represents an important element of good study design, particularly for complex interventions. Whole-diet or broader lifestyle interventions, including diet, may be considered 'complex interventions', and should be developed according to the UK Medical Research Council's guidelines on developing and evaluating complex interventions. Development of interventions according to these guidelines will involve the use of qualitative research methods (see Chapter 10).

Randomised controlled trials: Parallel and cross-over

Once these early studies have been completed, studies with greater rigour, in which participants are randomised to study groups, will test the hypothesis that the nutrient, food or diet will alter the selected outcome measures. Usually a series of studies will be conducted, with later studies extending the work as the evidence accrues. Examples include increasing the range of populations studied, using new and/or longer-term outcome measures, assessing the minimum effective amount (or 'dose') to be consumed, and evaluating different forms of presentation or delivery of the nutrient or food.

In any controlled study, in addition to measuring outcomes in participants receiving the active nutrient, food or dietary intervention, the same outcome measurements will be collected in a control group. The inclusion of a control group, which may receive either a placebo or no intervention, allows control outcomes to be compared with intervention outcomes and therefore increases confidence that changes observed during the study are directly attributable to the intervention. Without a control group, it is inappropriate to make cause-and-effect statements about an intervention, as other factors may be responsible for the effects observed. For example, if a study is conducted over several months without a control group, it is possible that any changes observed are attributable to normal seasonal changes rather than the intervention itself. As well as allowing seasonal variations to be taken into account, having a control group also means that the 'placebo effect' can be assessed. In some cases, just taking a supplement or eating in a different way is enough to make an individual feel 'better' to some extent, and this is particularly relevant when dealing with more subjective outcomes such as quality-of-life scales.

Two basic randomised controlled trial (RCT) study designs are encountered: *parallel group studies* and *cross-over studies*. The key features of these study designs are illustrated in Figure 3.1. In parallel group studies, each participant receives only one of the nutrition interventions (e.g. product A or B, or low intake or high intake) under study. Comparisons between groups must therefore be made on a between-participant basis. However, in some studies it may be feasible to use a different design in which participants receive more than one intervention. In cross-over studies, participants receive all interventions under comparison and the design specifies the order of interventions. This has the advantage that comparisons between interventions can be made on a within-participant basis, with a consequent improvement in the precision of comparisons and therefore in the power of the study, and a reduction in the required sample size. In such designs, participants act as their own controls. In a

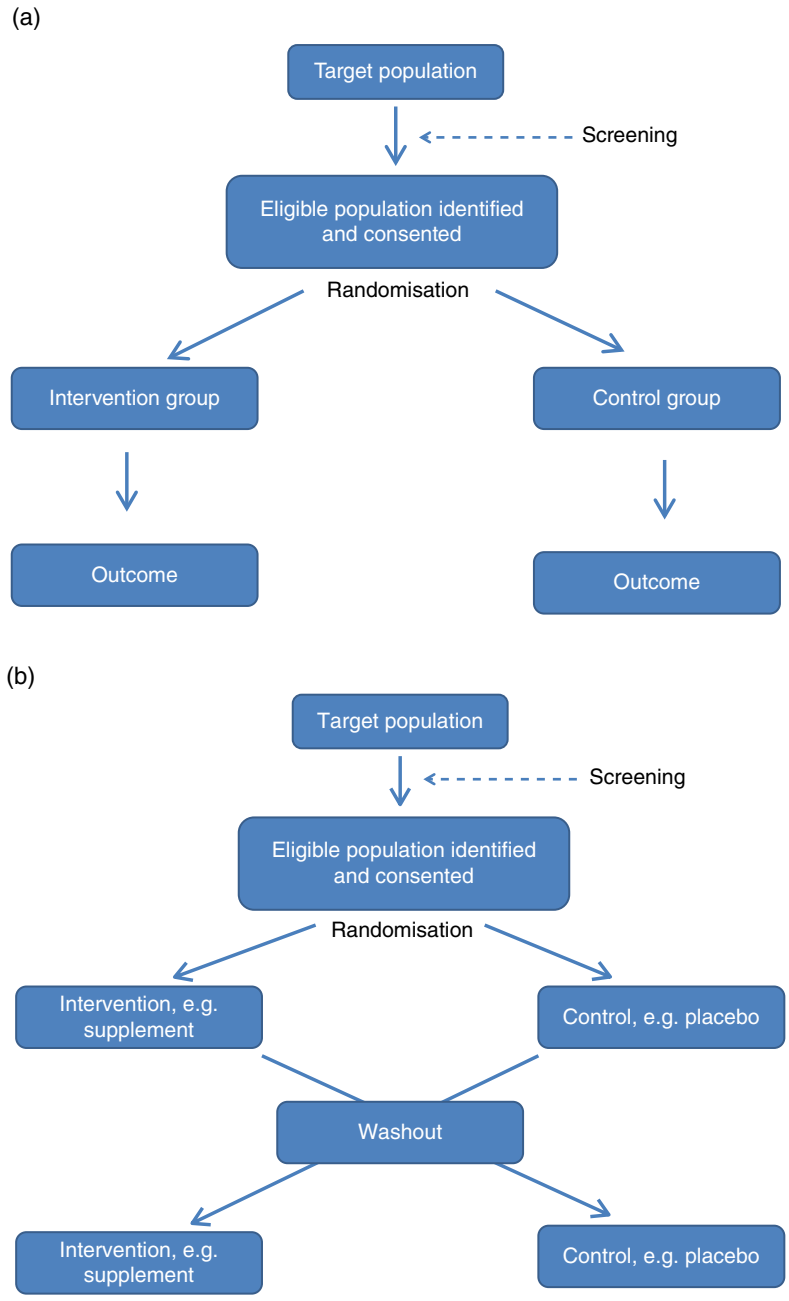

Figure 3.1 Study design for parallel and cross-over RCTs. Acute studies can follow either of these general designs, but durations will be markedly shorter. (a) Parallel group randomised controlled trial flowchart; (b) Cross-over randomised controlled trial flowchart.

cross-over design for two interventions, the participants are allocated to two groups that receive interventions in a different order. Assessments are performed at the end of each intervention period, although in some cross-over studies baseline measurements may also be taken at the start of each intervention period. Depending on the intervention and outcome measure, a washout period may be required between intervention periods to avoid contamination or carry-over effects; that is, the effect of an intervention given in one treatment period extending into the following treatment period(s). A run-in period may also be desirable in advance so as to minimise order effects. During this period participants may be asked to avoid certain foods. A Latin square design may be used, where appropriate, to extend cross-over studies to more than two interventions. However, since participants receive all interventions, increasing the number of interventions will extend the study duration and so may add to the participant drop-out rate.

For studies that require longer-term interventions, parallel studies are usually preferred, because of their shorter overall time frame. Furthermore, parallel studies are essential where a washout period may be ineffective at returning outcome measures to baseline, for example in certain tests of cognitive function. Parallel studies are also required where intentionally returning to baseline may be unethical, for example if body weight or bone mineral density is the outcome measures. Parallel studies are least suited to outcomes that show large inter-participant variation. Cross-over studies are favoured where participant availability may be restricted and in very short-term studies, for example postprandial studies to evaluate glycaemic responses or effects on satiety and short-term energy intakes. However, they are adversely affected by dropouts and necessitate a more complex analysis methodology.

The choice of study design will depend on these considerations, but also on the time frame, the availability of other resources such as cost, the level of financial support and research staff time available, and the potential roles of confounding factors, such as seasonal variations. Using cost as an example to be considered, the sample size for cross-over studies will be smaller, and the time frame for recruitment therefore shortened, but the overall time frame will be longer than for a parallel trial, as participants need to complete both intervention and control arms with an appropriate washout period. The effect of these differences on cost would have to be estimated for each individual study.

Other less commonly used types of RCT include the *factorial design* (in which all possible combinations of two or more interventions are tested, therefore permitting the evaluation of intervention interactions) and the *cluster randomised design* (in which the unit of randomisation is not the individual but a cluster of individuals defined, for example, by family, school class or primary care group). Further guidance on these designs is available in statistical texts on clinical trials.

Quasi-experimental studies

Like RCTs, quasi-experimental studies are designed to estimate the impact of an intervention on a group of participants. Although they can be similar to RCTs in design, they lack one or more key features of a true experiment; most commonly the element of random assignment to the intervention or control group is absent and sometimes the control group is lacking altogether.

Quasi-experimental studies are often used in public health, for example community food-based interventions. They are easier, quicker and cheaper to implement than RCTs and so require less forward planning and a shorter lead-in time. In some situations quasi-experimental studies are the only viable option, as it may not be ethical to have a control group that does not receive any intervention, for example in the provision of vitamins to infants. Some public health interventions, by necessity or for practical reasons, need to be rolled out quickly and on a wide scale, which excludes the incorporation of a control group. Quasi-experimental studies can provide valuable information about the potential usefulness of an intervention, but their internal validity (i.e. their ability to establish causality) will be compromised compared to the RCT design and this limitation must be appreciated when interpreting their results.

There are many different variations of quasi-experimental studies, but two frequently encountered designs are:

- *A before-and-after study without a control group*. In this case, data are collected on the endpoint of interest before and after an intervention takes place, but there is no control group for comparison and so it is not possible to be certain that any differences that have occurred between the start and end of the study are directly attributable to the intervention. It is possible that something else happened between the before and after measurements that influenced the results, or it is possible that completion of the pre-test assessments influenced completion of the post-test assessments, if non-objective data-collection methods such as questionnaires were used. For example, an intensive education intervention to reduce fat intake in a group of overweight participants took place over a six-month period and, at the same time, a public health campaign targeting fat intake was launched. Without a control group it would not be possible to say whether changes in fat intake that took place during the study were attributable to the intensive education intervention or to the public health campaign.

- *A before-and-after study with a non-equivalent groups design*. In this type of study, one group of individuals is recruited and assigned to the intervention and another group of participants is chosen to act as the control group. Since the groups are not created through random assignment, they may not be similar (or equivalent) in all key aspects at the start of the study and this may affect the outcome of the study and thus its internal validity; that is, its ability to conclude that the intervention was causally related to the study outcome. For example, in a study examining the effects of a dietary intervention on total cholesterol, if participants are not randomly assigned to the intervention or the control group at study outset, it is possible that participants in the intervention group may have lower total cholesterol concentrations and a healthier diet at the start of the study compared to the control group. In this case, the intervention group would be unlikely to benefit as much from the intervention as the control group, thus the non-equivalence of the groups at the start of the study would bias the results towards the null hypothesis.

Population-based fortification studies

Food fortification is defined by the World Health Organization (WHO) as 'the practice of deliberately increasing the content of an essential micronutrient, i.e. vitamins and minerals (including trace elements) in a food, irrespective of whether the nutrients were originally in the food before processing or not, so as to improve the nutritional quality of the food supply and to provide a public health benefit with minimal risk to health'.

Food manufacturers can fortify foods on a voluntary basis, in line with government legislation, meaning that the individual can make a choice about whether to purchase such foods or not. An example of this would be fortification of ready-to-eat cereals. In contrast, population-based food-fortification programmes are sometimes implemented as part of public health policy to correct dietary deficiencies (e.g. iodised salt to prevent iodine deficiency) or enhance the status of a micronutrient to a level that will prevent specific undesirable health outcomes (e.g. fortification of flour with folic acid to prevent neural tube defects). Population-based fortification programmes require careful planning and consideration of a wide range of background scientific data before commencement, including the following:

- Examination of high-quality data on the dietary intake (usual food intake and dietary patterns) and nutritional status of the population, including age- and sex-specific subgroups, in order to inform decisions about the most appropriate food vehicle for fortification and to allow modelling of dietary exposure in relation to tolerable upper limits as part of the overall risk-assessment process.
- Calculation of the dose and most appropriate form of micronutrient to add based on: data from efficacy and effectiveness trials; the food vehicle chosen; the bioavailability of the nutrient in question when delivered in the food matrix; and findings from risk-assessment modelling.

Furthermore, careful monitoring for undesirable consequences (e.g. over-exposure in certain population subgroups resulting in toxic side effects) as well as desirable effects (improved population micronutrient status, reducing the incidence of the targeted adverse health outcome) of the fortification programme in the short, medium and long term is paramount and should be carefully planned before programme implementation.

3.3 Considerations when planning intervention studies

The major factors involved in the planning, conducting and reporting of intervention studies are identified in Table 3.1, which uses a similar structure to that in the Consolidated Standards of Reporting Trials (CONSORT) checklist for clinical trials. An expanded discussion of these factors appears in this section. These factors will be most relevant for RCTs, as described above, but some factors will also apply to intervention studies in general.

Table 3.2 gives some examples, from peer-reviewed journals, of the different study designs described in Section 3.2. Some published studies, particularly earlier ones, do not give clearly stated null hypotheses based on a single primary outcome measure. In these cases, the null hypotheses in Table 3.2 have been inferred from the hypotheses, aims or objectives given in the paper. The distinction between null hypotheses and alternate hypotheses is outlined later in this section. Achieving a study design that fully satisfies all the considerations described here may be constrained in practice by a number of factors, which include practical and logistical issues and the availability of resources, eligible participants and appropriate outcome measures. Thus, the purpose of Table 3.2 is to illustrate the range of types of study design that have been used, rather than to provide examples that may be considered to satisfy fully all the considerations described.

Hypothesis

The primary hypothesis, which is tested statistically, should be framed as a null hypothesis, which states that there is no difference between the tested intervention

Table 3.1 Factors to consider and recommended standards for human intervention trials evaluating health benefits of nutrients, foods and diets. Modified from Welch *et al.* (2011) and Woodside *et al.* (2013).

Phase	Factor	Recommended standard
Design	Hypothesis	Clear hypothesis
	Study design	Appropriate design, randomised where possible
	Duration	Appropriate to design, intervention and outcome measures
	Intervention	Test and control interventions suitably matched
	Amount	Appropriate to outcome measures and to practical usage
	Outcome assessment	Define primary outcome and method of measurement
		Define all secondary outcomes and methods of measurement
	Eligibility criteria	Define all eligibility criteria
	Statistical considerations	
	Randomisation	Use randomised design; ensure appropriate allocation, sequence generation and concealment
	Blinding	Ensure double blinding if feasible, single blinding if not
	Size of study	Conduct power calculation based on primary outcome measure
Conduct	Study protocol	
	Ethical approval and trial registration	Obtain approval, register trial, comply with Declaration of Helsinki
	Recruitment	Define recruitment strategy and process, including settings and dates
	Data collection	Define relevant measures, select suitable methods for assessment, collection and analysis
	– Demographics, lifestyle, background health status and diet, and diet changes	
	– Adverse events and unintended effects	Use suitable methods to record and respond appropriately
	Compliance	Define acceptable level, strive to maximise, assess
Analysis and interpretation	Statistical analysis	Devise appropriate analysis methods, based on study design and outcome measures
	Discussion and interpretation	Consider study limitations and generalisability of findings
	Conclusions	Relate directly to hypothesis, study design, intervention and participants

and the control (see Table 3.2). If the statistical test rejects the null hypothesis, then the alternative hypothesis is accepted, indicating that there is a difference between the two interventions. The primary hypothesis to be tested directly influences all aspects of the study, including the study design and duration, the eligibility criteria, the amount of food or nutrient that will be provided and the nature of the control group. The hypothesis should be based on a thorough review of the available evidence. This review should not only encompass other intervention studies, but also consider epidemiological, animal and *in vitro* studies. Where possible, all available evidence should be reviewed systematically and an assessment of safety and potential risks should be carried out. The primary outcome measure should be clearly defined and must relate to the primary hypothesis.

Duration

The study duration must be long enough to allow changes in the primary outcome measure and will be determined by data from previous intervention studies and from knowledge of the underlying physiology and biochemistry, for instance relevant tissue-turnover rates. The duration must also relate to the timescale of the hypothesis, which may address acute effects (e.g. glycaemic response or increased alertness) or longer-term outcomes. Thus, no standard can be set for duration, but the aim should be to set the shortest feasible duration for ethical reasons, to conserve resources and to avoid participant fatigue leading to non-compliance or withdrawal. In some cases, post-study follow-up measures are desirable to evaluate persistence or other longer-term effects, although such follow-up can add significantly to study costs.

Intervention nutrient, food or diet

The intervention will be the nutrient, food or diet under investigation. Consideration must be given, however, to the intended use of the intervention, and the study design should take this into account. For example, if it is intended that a food should be consumed as part of a mixed meal, once a day, then the study design should be testing that pattern of consumption and details of frequency and timing of ingestion reported. If a particular food is the

Table 3.2 Examples of the different study designs that have been used in human nutrition intervention studies published in peer-reviewed journals.

Null hypothesis	Duration	Intervention	Control	Main outcome measures	Participant and eligibility criteria	Randomisation	Blinding	Source
Exploratory, feasibility or pilot studies								
Phytoestrogen supplementation does not increase urinary phytoestrogen metabolite excretion or a range of biochemical biomarkers	One day and one week	80 mg mixed phytoestrogen supplement per day	None	Phytoestrogen metabolites, lipids, antioxidant status, DNA damage, insulin status	10 healthy women	No	No	Woodside *et al.* (2006)
A low-carbohydrate, ketogenic diet has no effect on glycaemia and medication use in free-living overweight and obese patients with type 2 diabetes	16 weeks	Low-carbohydrate, ketogenic diet	None; before vs after intervention	Haemoglobin A1c, glucose, insulin, medication use	28 men and women with Type 2 diabetes, 35–75 years; body mass index (BMI) $> 25\,kg/m^2$	No	No	Yancy *et al.* (2005)
A self-monitoring weight-management smartphone app is neither acceptable nor feasible as a standalone weight loss intervention	6 months	Smartphone self-monitoring app	No control per se; similar self-monitoring information to smartphone app group delivered via a website or paper diary	Feasibility and acceptability (adherence to the trial and adherence to the intervention); secondary measures – anthropometry	128 overweight (BMI $> 27\,kg/m^2$) men and women aged 18–65 years	Yes	Researchers carrying out anthropometry assessments blinded to group allocation	Carter *et al.* (2013)
Modified citrus pectin (MCP) does not increase prostate-specific antigen doubling time	One year	14.4 g MCP taken as 6 capsules 3 times per day with water or juice	None; before vs after intervention	Prostate-specific antigen doubling time (PSADT)	Men with prostate cancer after localised treatment (n = 13)	No; non-randomised	No	Guess *et al.* (2006)

Randomised controlled trials – parallel design								
Consumption of wheat aleurone does not increase plasma betaine or affect related biomarkers	4 weeks	Ready-to-eat cereals and bread providing 27 g wheat aleurone/day	Ready-to-eat cereals and bread balanced for macronutrients and fibre	Plasma betaine, choline, B vitamins, homocysteine, LDL-cholesterol	80 men and women at risk of metabolic syndrome (45–65 years; BMI ≥ 25 kg/m^2)	Yes	Single (participant blinded)	Price *et al.* (2010)
Consumption of nuts does not increase antioxidant status in participants who may have impaired antioxidant status	8 weeks	Habitual diet + cashew or walnuts (20% of daily energy intake)	Prudent control diet; energy adjusted to maintain body weight	Antioxidant status	68 men and women with diagnosed metabolic syndrome according to NCEP ATP III criteria	Yes	No	Davis *et al.* (2007)
Consumption of whey peptides with *in vitro* ACE-inhibitory properties does not decrease blood pressure	12 weeks	125 ml milk drink per day with whey peptides	125 ml milk drink per day without whey peptides	Blood pressure, selected inflammatory markers, insulin, glucose	54 patients with mild hypertension not receiving ACE inhibitors or angiotensin II receptor blockers	Yes	Double	Lee *et al.* (2007)
Probiotics do not prevent gestational diabetes in high-risk pregnant women	16 weeks' gestation to delivery	Probiotic capsule – 1 x 10^9 cfu each of Lactobacillus rhamnosus GG and Bifidobacterium lactis BB-12 per capsule	Placebo capsule	Diagnosis of gestational diabetes	540 women recruited at 14–16 weeks' gestation (singleton pregnancy) with BMI > 25.0 kg/m^2	Yes	Double	Dekker Nitert *et al.* (2013)
Isoflavone-enriched foods do not affect bone mineral density or hormone status	One year	110 mg isoflavones per day in biscuits and bars	Isoflavone-free biscuits and bars identical in composition, taste and appearance	Bone mineral density, panel of hormones, bone biomarkers, lipids and routine clinical chemistry profile	300 women, Caucasian, menopausal for 12–60 months, non-osteoporotic BMI 22–29 kg/m^2	Yes	Double	Brink *et al.* (2008)
Beta-carotene supplementation does not reduce risk of malignant neoplasms and cardiovascular disease	12 years	50 mg beta-carotene supplement in capsules on alternate days	Placebo capsules on alternate days	Malignant neoplasms, cardiovascular disease incidence or overall mortality	22 071 male physicians, 40–84 years; 11% current and 39% former smokers	Yes	Double	Hennekens *et al.* (1996)

(*Continued*)

Table 3.2 (*Continued*)

Null hypothesis	Duration	Intervention	Control	Main outcome measures	Participant and eligibility criteria	Randomisation	Blinding	Source
The Mediterranean Diet has no effect on primary cardiovascular disease prevention	Median follow-up 4.8 years	Dietitian-led advice to follow a Mediterranean diet plus provision of key foods (either 1 l extra-virgin olive oil/week or 30 g mixed nuts/day)	Control diet (advice to reduce dietary fat; similar frequency and intensity of dietary advice as intervention groups)	Primary endpoint: composite of myocardial infarction, stroke and death from CVD	7447 men (aged 55–80 years) and women (aged 60–80 years) with no CVD, but who either had diabetes or at least three CVD risk factors	Yes	No (endpoint confirmation conducted blind)	Estruch *et al.* (2013)
Randomised controlled trials – cross-over design								
Food and energy intake during ad libitum ingestion of pasta, rice or potato with a meal does not affect food and energy intakes, or insulin and ghrelin levels	One day	Ad libitum intake of meals with pasta, rice or potatoes.	No control per se; a comparison of potatoes, rice or pasta	Food intake (g), energy intake, satiety and hunger feelings, blood insulin, ghrelin and glucose	11 participants with no signs and symptoms of an acute or chronic disease or taking medication; no family history of diabetes mellitus	Yes	No	Erdmann *et al.* (2007)
Availability of different-sized food portions does not affect food and energy intake sensations in normal-weight and overweight adults over four consecutive days under fully residential conditions	4 days	Provision of standard or large portions of the same foods and beverages in a residential setting	No control per se; a comparison of standard and large portions	Food intake (g) and energy intake	44 men and women; 18–65 years; BMI 18·5–30 kg/m^2; non-smokers; omnivores, apparently healthy	Yes	Single (participant blinded)	Kelly *et al.* (2009)
Cocoa flavanols from cocoa do not affect dermal microcirculation	One day	100 ml cocoa drink with high flavanol content	100 ml cocoa drink with low flavanol content	Skin microcirculation – blood flow and velocity measured by echo Doppler	10 healthy women, non-smoking and non sunbathing	Yes	Not reported	Neukam *et al.* (2007)

Quasi-experimental studies								
Before and after study								
Cooking classes will not improve the dietary intake of people with type 2 diabetes	Four weekly cooking classes – each 3 hours' duration	Series of cooking classes for people with type 2 diabetes and their family members held in community locations including schools, churches and senior centres	No	Change in dietary intake (energy, macronutrients, sodium) assessed using three-day food records completed prior to attending cooking school and one month after completing the classes	Type 2 diabetes	n/a – no control group	No	Archuleta *et al.* (2012)
Non-equivalent groups design								
Provision of a new food hypermarket in a 'food-retail deficit' community in Glasgow will not increase food availability or have a positive effect on fruit and vegetable consumption, self-reported or psychological health	10 months	Natural experiment – opening of a new food hypermarket in one area of Glasgow	A matched 'comparison' community in Glasgow	Before and after questionnaire assessment of fruit and vegetable consumption, self-reported and psychological health; data for the intervention community compared with data for the 'comparison' community	Intervention community and 'comparison' community were 5 km apart; postal questionnaires distributed to a random selection of households (by postcode) in each community two months before the new hypermarket opened and again 10 months after	No	No	Cummins *et al.* (2005)

intervention being tested, then investigators need to decide whether participants will substitute the test food for habitual foods, whether the test foods will be added to their usual diets, or whether some sort of food-exchange model can be implemented with participants, as each of these scenarios will be answering a slightly different research question. This section outlines some factors to consider when planning the intervention.

Amount consumed

The dose of a nutrient or other component, or the amount of the food to be consumed, will depend on a number of factors (e.g. previous data, underlying physiology, food matrix, palatability and bioavailability). However, the amount to be consumed should be close to that intended for practical use. Furthermore, it is important to test and document the amount of the nutrient or food that is provided, for example by directly measuring the amount of a particular nutrient present in a supplement capsule.

Control group intervention

The control is a food, nutrient, substance or product that does not provide the component that is being tested, and its composition should also be analytically documented. The control should be matched for sensory characteristics and taken in the same way as the test intervention. A control is relatively easy to achieve in supplementation studies using pills or similar preparations by producing a placebo preparation. However, in studies of foods or whole diets, it is more difficult, and perhaps impossible, to develop a control intervention identical to the test intervention but not containing the active component(s) under study. Blinding may not be possible for many foods where the intervention is easily identifiable by both trial participants and researchers, as may be the case with some minimally processed foods such as fruit or vegetables, and some manufactured consumer foods such as cereal products. However, some degree of blinding may be made possible by the use of suitable packaging that conceals products from the researchers and study participants. If the aim is to use a single food group, such as fruit or nuts, then the formulation of a control food is impossible and instead the control arm would receive either no food or a smaller number of portions of the food being studied; this may have effects on other aspects of diet and behaviour. For whole-diet interventions, for example the Mediterranean Diet, it is usual to measure self-reported adherence to that diet using a previously developed scoring scheme, with control groups not receiving the dietary advice and therefore being less adherent to the whole-diet pattern and consequently attaining lower scores. Further guidance on attaining an ideal control is available in other published literature, but is likely to vary depending on the type of intervention being tested.

Outcome measures

All intervention studies will assess outcome measures and will compare these between intervention and control groups, if a control group features in the study design. Most studies will have a range of outcome measures, but the study should be powered based on the pre-specified primary outcome measure, as stated in the hypothesis, and the sample size calculated using that outcome measure (see the discussion of size of study later in this chapter). Similarly, if an outcome is assessed at several time points over the course of the study, either a single time point or a single summary measure of results at several time points should be pre-specified as the primary outcome measure. All outcome measures, whether primary or secondary, should be stated and defined in the study protocol.

It is essential that the outcome measure is of biological relevance. In some cases the outcome measure is clearly relevant, as it is a direct, objective measure of the impact on nutritional intake or status (e.g. energy intake or nutrient concentration in plasma – see Chapters 4, 6, 11 and 12) or intended health effect (e.g. body weight, or diagnosis of a disease or muscle strength). Subjective measures are also used, such as feelings of health, appetite or fatigue; in these cases, it is important to use validated instruments if these are available. When the effect cannot be measured directly, indirect or surrogate factors such as biological markers or risk factors are used to reflect a functional, physiological or biochemical characteristic associated with a disease, or as a predictor of the later development of the disease. Examples include glycated haemoglobin as an indicator of long-term hyperglycaemia and risk of type 2 diabetes complications, plasma LDL-cholesterol as a measure of cardiovascular disease risk, bone mineral density as a measure of osteoporosis risk, complex metabolomic or proteomic profiles as markers of function and disease risk, and the presence of adenomatous colon polyps as an early indicator of colon cancer. Most indirect outcome measures are chosen because they reflect consensus guidelines or are commonly used by experts in the area. For example, detailed guidelines have been proposed for particular outcomes such as the assessment of glycaemic responses or satiety. However, very few markers have been assessed and validated by expert consensus in terms of their specificity, variability, limitations and applicability to a range of population groups.

Methodological aspects

An effort should be made to standardise all outcome measure assessments and reduce measurement error as far as possible (e.g. by standardising measurement protocols, training observers and averaging several measurements rather than using a single measurement), especially if measurement errors are known to be large. Where possible, the

researcher assessing study outcomes should be blinded to the intervention assignment.

Analytical variability

Laboratory analytical methods should be precise, accurate, sensitive and specific, and these performance characteristics should be recorded in a file of standard operating procedures (SOPs) or similar-quality record documents for the study. Intra-laboratory analytical variability should be minimised by using automated equipment to analyse samples in duplicate or triplicate, in batches that represent the range of interventions, participants and sampling times, with suitable internal and external standards and participation in quality assurance programmes. Ideally, all samples from a study should be analysed at the same time, and all samples from an individual participant in one run, but this may be precluded by degradation in storage, even at low temperatures. Biomarkers that have high methodological variability will often require a larger number of trial participants to give the study adequate power.

Biological variability

Biological variability arises from many factors (e.g. genetic background, circadian rhythm, seasonal differences, menstrual cycle) and may introduce systematic bias. Thus, it is important to understand the factors underlying this variability for the biomarkers, and to take samples or adapt the study design accordingly.

Biologically meaningful changes

Although a trial may find a statistically significant change in an outcome measure, such a response does not necessarily mean that the intervention will be effective in terms of producing a discernible health benefit or risk reduction in the target group. Thus, the size of the change and its potential biological, clinical or public health significance should also be considered when performing the sample size calculation (see later in this chapter).

Selection of participants: Eligibility criteria

Eligibility criteria, which often include age, gender, health and disease status, are functional, physiological or clinical characteristics or demographic variables used to define the study population. Eligibility criteria may also include lifestyle factors, such as smoking habit or level of physical activity, and dietary factors such as low fibre intake or the consumption of restricted diets. Eligibility criteria can be presented as inclusion and exclusion criteria.

Eligibility criteria should describe participants adequately, so that the results can be appropriately interpreted in terms of their generalisability. Eligibility criteria should also be selected with the target population for the test intervention, as well as the hypothesis and outcome measures, in mind. Inter-participant variation may be reduced by using stricter eligibility criteria to select a more homogenous group of participants for the study. However, this approach also has the disadvantage of restricting the target population and consequently will limit the generalisability of the findings. Children and women of childbearing age will need to be excluded from any studies that may have an adverse effect on normal growth and development or have teratogenic potential.

It is important to define eligibility criteria using objective quantitative descriptors wherever possible. For example, many nutrition interventions use 'apparently healthy' participants. Health may be evaluated by using a questionnaire on medical history and surgical events, or this may be extended to a physical examination and screening of blood and urine. 'Health' may merely refer to the absence of diagnosed disease, or to a specific aspect such as a healthy blood pressure, and in such cases the criteria can be very specific and may follow official guidelines. However, 'apparently healthy' may also include a healthy lifestyle, which could be assessed using questionnaires, for example for physical activity, dietary habits, smoking, alcohol and medication use.

Statistical considerations

Randomisation

Randomisation is the allocation of participants to interventions using a random process such as the toss of a coin. It ensures that the investigator does not bias the study outcome by influencing the intervention to which a participant is allocated. The main advantage of random allocation is that it will produce study groups that are comparable with respect to both known and unknown factors that could influence the outcome measure. That is, it ensures that potential confounding factors are equally distributed between groups. Consequently, it increases the internal validity of the study, meaning that any observed difference in the responses of the two intervention groups is likely to be due to the effects of the intervention. Randomisation helps to ensure that the comparison of interventions is fair (by eliminating selection bias) and that the statistical analysis is valid.

To allocate individual participants to intervention groups, random number generation (either from tables or more usually by computer) is often used. However, it is advisable to ensure that approximately equal numbers of participants are assigned to each group by using a restricted (or block) randomisation, in which participants are divided into blocks within which equal numbers of allocations are made to each intervention. To avoid any possible predictability of the allocations at the end of a block, it is advisable to vary the block size. It is often desirable to stratify participants into subgroups

defined by important variables such as age, gender and ethnicity that could influence the response to intervention. A restricted randomisation is then conducted within each subgroup. Stratification will generally result in more comparable study groups and can also reduce variability in the response measure when incorporated into the statistical analysis. Minimisation, a technique that minimises imbalance between the participants in the intervention groups over a number of variables simultaneously, may offer a more practical approach than stratification on multiple variables.

Concealment of the intervention allocations

CONSORT highlights the importance of detailing who generated the study randomisation schedule, who allocated participants and what steps were taken to conceal the allocation in order to minimise bias, subconscious or otherwise. Successful randomisation should result in an unpredictable allocation sequence (i.e. the researcher will not be able to predict to which group the next participants will be assigned) and adequate concealment of the allocation sequences until the participant is made aware of their group assignment. In a multicentre trial, a telephone randomisation procedure can be implemented to safeguard the allocation sequence. For a small, single-centre trial, a simple way to eliminate any possible bias of this sort is to implement randomisation using sealed envelopes. In this process, the random intervention allocations are concealed in sequentially numbered, opaque, sealed envelopes, prepared by a researcher who is not involved in the recruitment or allocation of participants. Only after a participant has given consent, been enrolled in the study and the envelope endorsed with the participant's name should the seal be broken to reveal to which intervention the participant has been allocated. This process ensures that knowledge of forthcoming assignments is not available to researchers and shields the allocation sequence until assignment occurs.

Blinding

The assessment of study outcomes may be influenced by knowledge of which intervention was received, particularly for subjective outcomes. Such bias can be avoided by using blinded assessment. If neither assessor nor participant knows which intervention the participant received, then the study is double blind. If the participant knows but the assessor does not (or vice versa), then the study is single blind. Blinding should also be carried through into laboratory determinations and statistical analysis. The time of unblinding, which is usually after the freezing of the database (i.e. when all data entry for the study is completed and the study database has been checked and finalised), should be documented in the study report and may be mentioned in any subsequent document.

Where possible, and particularly for food products, the effectiveness of blinding should be assessed at the end of the study and commented on in the study report. This can be achieved by the use of a simple questionnaire asking participants which product (test or control) they thought they were consuming.

Size of study (power calculation)

It is essential to estimate the number of participants required for the study. A study that is too small is likely to fail to detect important differences between interventions, while one that is too large may needlessly waste resources and would be unethical. In certain circumstances trials may be designed to be analysed after every participant's result becomes available (sequential design) or after pre-specified numbers of participants' results become available (group sequential designs). These designs are ethically appealing because they ensure that inferior interventions are quickly identified, so minimising the numbers receiving them. However, even when such early termination is feasible it is not always advisable, since it can lead to intervention effects being estimated with poor precision.

The usual methods for sample size estimation require specification of the magnitude of the smallest meaningful difference in the outcome variable. The study must be sufficiently large to have acceptable power to detect this difference as statistically significant, and must take into account possible non-compliance and the anticipated drop-out rate. Information about the degree of variability in the outcome is also required and may come from previous published or unpublished data, or from a pilot or exploratory study specifically performed for the purpose (discussed earlier in this chapter). A multicentre study may be necessary if the study is too large to be performed in a single centre. Statisticians are key members of research teams and it is recommended they are involved at an early stage, not only in study size calculation but also in planning the design and analysis of the study.

Ethical approval and study registration

Researchers should determine the appropriate local ethical approval and research governance procedures required for their study, and seek these approvals before the study commences. While not all nutrition research may be classified as medical research, it is recommended that researchers adhere to the World Medical Association's Helsinki Declaration. One of its recommendations is that every clinical trial (including human nutrition intervention studies) must be registered in a publicly accessible database before recruitment of the first participant. Such registration, with accompanying protocol details, is

intended to reduce the consequences of non-publication of studies (for example, repetition of negative studies), of selective reporting of outcomes and of reporting per protocol (PP) rather than intention to treat (ITT) analyses (see on the discussion of statistical analysis later I this chapter). The WHO has stated that 'the registration of all interventional trials is a scientific, ethical and moral responsibility', while the International Committee of Medical Journal Editors only considers trials for publication if they are registered before enrolment of their first participant. The academic view is that a priori trial registration is essential for ethical research in humans.

Recruitment and participant flow

The study protocol should state the methods by which participants will be recruited, and details of the recruitment process should be carefully described, with details of numbers of participants approached, screened, recruited and completing, and reasons noted for non-recruitment (ineligibility, lack of willingness to participate) and non-completion. Informed consent should be obtained. When reporting the study, this information is best summarised in a participant flow diagram, such as that suggested by CONSORT (as illustrated in Figure 3.2).

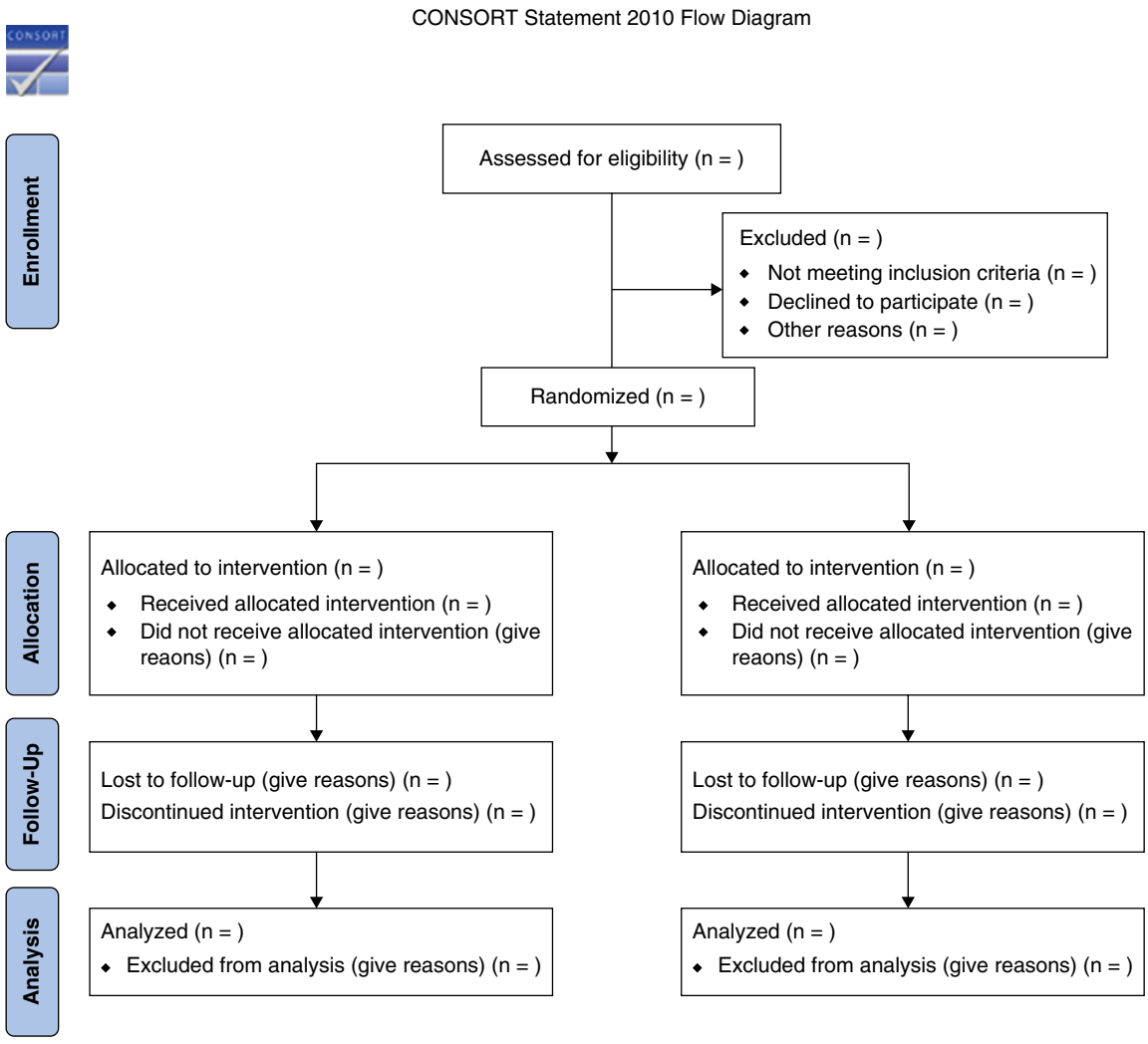

Figure 3.2 Flow diagram of the progress through the phases of a parallel randomised trial of two groups. Schulz, K.F., Altman, D.G. and Moher, D. (2010) CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, **340**, c332.

Data collection

Data should be collected using a standardised case report form. Participants should be assigned a unique study number at the start of the study, and all their data should then be held under that study number. That is, no participant-identifiable information should be held by the researchers, other than a single sheet where the study number is linked to the participant contact details. All data, both paper and computer-based, should be kept securely and all data collection conducted in line with the required local ethical and research governance regulations.

Background diet and change in diet during intervention

The nature of the participants' background diet may be one of the eligibility criteria. Regardless of this, and particularly in longer-term studies, it is important to collect background dietary information in order to characterise the participants' habitual diet in terms of nutrient intake, food consumption and overall dietary pattern. Diet should also be assessed during longer-term interventions in order to detect changes in it over time that may potentially confound the results of the study. Such an assessment will be particularly important when the intervention is with whole foods or whole diets, where the control arm is more difficult to design and define, and where blinding is not possible. Where a nutritional supplement is being tested against a placebo, randomisation has been performed and double blinding has been possible, any dietary changes over the course of the intervention period would be expected to be equally distributed between the intervention and control groups. However, with food or whole-diet interventions where participant blinding is impossible, dietary changes will differ between intervention and control groups, and full dietary assessment is particularly important to establish, for example, how a particular food or food group has been incorporated into the overall diet; whether other foods have been displaced as a result of the intervention; and the impact that has had on overall diet quality and nutrient intake. For some outcome measures that are affected by body weight, such as insulin resistance, assessment of the impact of a dietary change ideally requires body weight to be maintained over the course of the intervention, and therefore intervention and control diets will have to be carefully energy matched and weight monitored during the intervention period.

A number of dietary assessment methodologies are available, including retrospective tools such as a food frequency questionnaire or diet history, and prospective methods such as a food diary or weighed food record (see Chapter 4). However, dietary intake assessment methods are subject to misreporting. In order to check the reliability of dietary data, reported energy intakes should be compared with the estimated energy requirement for each participant and compared to established cut-offs for under- or over-estimating energy intake. This is particularly important if these assessments are being used as a way of monitoring compliance.

Background health status and lifestyle, and changes in health status and lifestyle during intervention

In addition to their possible role as eligibility criteria, it is also important to characterise the study population in terms of demographic background, health status and lifestyle behaviours, in order to allow appropriate interpretation and generalisation of the results. Examples include age, gender, level of medication use, years of formal education, socio-economic status, physical activity and smoking habit. The monitoring of health status and lifestyle behaviours should also be carried out in the course of longer-term studies to assess potential between-group differences, which may confound outcome measures.

Adverse events

An adverse event (AE) is any unfavourable and unintended sign (including an abnormal laboratory finding), symptom or disease temporally associated with the use of an intervention, whether or not it is considered to be related to the intervention. Recording AEs is of major importance in pharmaceutical studies, allowing a risk–benefit analysis. Hence, there is an abundance of guidelines for the management of AEs in the clinical study setting (e.g. European Medicines Agency; International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; US Department of Health and Human Services, Food and Drug Administration). There are no guidelines for nutrition intervention studies, given that these studies involve testing foods, supplements or ingredients in participants that are usually apparently healthy. However, the formal recording of AEs is required for good practice in nutrition research.

It is generally regarded as good practice to record all AEs, no matter how trivial, in a participant's file. In the case of nutrition studies, AEs are likely to be very minor in nature, for example mild nausea or minor gastrointestinal discomfort. Such occurrences may be the result of changes in dietary pattern or consumption of unfamiliar products and will often lessen over time as the body adjusts to the dietary changes. These minor events are sometimes known as unintended effects (to use recent CONSORT terminology). Their recording is desirable and important in human nutrition interventions, as it contributes to data on the tolerability of the product.

Some of these minor occurrences will be anticipated by investigators; if so, questionnaires should be used to provide quantifiable data, employing standardised formats where available, for example to assess gastrointestinal effects such as bloating or flatulence. Data should be collected at baseline and at suitable intervals during the study to assess onset and time course. Time, intervention and group effects should be tested statistically and, if significant, potential influence on compliance, withdrawal and outcome measures should be considered.

Any serious or unexpected adverse events that are encountered, whether or not they appear to be related to the intervention, should be reported immediately to the lead researcher, the relevant research ethics committee, the sponsor and other relevant regulatory bodies for review and appropriate management.

Compliance

Any deviations from protocol can affect the validity and relevance of an intervention study. Low levels of participant compliance in nutrition studies decrease the power to detect effects on specified endpoints, result in false negative findings, and ultimately mean that the study is unable to provide evidence to support or refute a potentially beneficial effect of the intervention. Poor compliance in a particular subgroup will also reduce the generalisability of the results and has implications for wide-scale implementation of the intervention. When compliance is very different between allocated groups, this may be because acceptability of the interventions differs. Therefore, a nutrition intervention study should aim to have measures in place to maximise and assess compliance.

Methods to encourage and measure compliance

The choice of compliance assessment methods will depend on study design, duration and intervention type. In acute or postprandial studies, the intervention is usually consumed only once, or a limited number of times, under supervision, and thus compliance is not usually an issue. However, maintaining compliance throughout longer-term studies is very important and may employ one of the strategies discussed here. Consumption under supervision throughout the dietary intervention will maximise compliance; however, this has resource implications, as it will require the use of a special nutrition facility and an intensive level of research staffing (observations may last a few hours, be at meal times only or extend to residential studies lasting several days or weeks). The complete provision of intervention supplements or food or diets for consumption in a free-living situation is a more commonly used approach and participants would be asked to return any unconsumed items. However, in this case an assumption is made that all unreturned items have been consumed, which may not be the case. In addition to providing intervention foods, maintaining regular contact with participants is key to achieving good compliance, as it allows any issues to be identified and dealt with at an early stage. Furthermore, informing participants that compliance will be measured is likely, in itself, to improve adherence to the dietary intervention. Dietary records, such as food diaries or diet recall methods, can be used to measure compliance, but such self-reported intake data are predisposed to errors (see earlier in this chapter). Thus, the assessment of tissue biomarkers as independent and objective measures of compliance is preferred when possible (e.g. serum selenium or fatty acid composition of erythrocyte membranes; see Chapter 8).

Acceptable levels of compliance

Acceptable levels of compliance for human nutrition studies have rarely been stated and are difficult to comment on definitively (see later in this chapter for discussion of how compliance will affect statistical analysis). A decision about the statistical analysis approach will be partly influenced by whether studies are designed as tests of efficacy (biological effect) or effectiveness (with the potential to modify outcome in a real-life situation), as the former studies will be more focused on maximising compliance. Making a decision on an acceptable level of compliance relies on an accurate, objective assessment of compliance being available. A priori decisions should be made regarding the acceptable level of compliance for inclusion in a PP analysis. For example, in a supplement study a level of consuming more than 80% of the supplements provided might be specified as an indicator of good compliance.

Statistical analysis

There are a number of statistics books that cover the basics of randomised intervention trial methodology, both in the design and analysis phases. It is good practice to include a statistical analysis plan that specifies the statistical methods to be used in the trial protocol. The hypotheses to be tested for both primary and secondary outcomes (including whether they are one-sided or two-sided) and the significance level to be employed should be clearly stated.

Rationale for using statistical methodology

In common with other research in medicine and the biological sciences, the differences between groups that the investigator wishes to identify in a nutrition study are usually masked by several types of variation (inter- and

intra-participant variation, measurement error and so on); strategies to minimise these have been outlined earlier in this chapter.

These errors mean that there is a need for the results of a study to be assessed objectively using appropriate statistical methodology. This section describes the basic statistical concepts necessary for the analysis of nutrition intervention studies. Although tests of hypotheses play a key role here, it is worth emphasising that the calculation of confidence intervals for intervention effects can often be more informative.

In general, statistical techniques require an assumption that the group under study may be considered to be a random sample from a target population about which inferences are to be made. In practice, there would be considerable practical difficulties in mounting an intervention study on a truly random sample from a target population, and usually a convenience sample such as a group of healthy volunteers or patients attending a hospital out-patient clinic will be studied. The investigator should be particularly cautious in any extrapolation of findings beyond the population from which the study sample was drawn. It is also worth emphasising that statistical methods will only take account of sampling error (i.e. variation arising from the process of sampling); they cannot quantify the extent of biases attributable to non-random sampling, particularly bias that may be introduced through losses to follow-up.

Preliminary steps in data analysis

Before attempting any formal statistical comparisons, it is important to visualise the data with histograms and scatter diagrams to examine the shapes of distributions, to check for outliers and to establish the nature of any relationships between variables.

Suitable descriptive statistics should also be presented to characterise the participants under study, and an indispensable step is to construct a table of participant characteristics by group. For quantitative variables, this should include both measures of location and measures of dispersion, typically the mean and standard deviation for roughly symmetrically distributed variables or the median and interquartile range for variables whose distribution is heavily skewed. For categorical variables, both frequencies and percentages should be included in this table. In an adequately randomised study it is not usually considered necessary to perform statistical tests on these baseline group characteristics, since any differences observed between groups must be due to chance.

Hypothesis tests for comparing groups

Along with the study design, the scale of measurement of the response variable is of fundamental importance in deciding which statistical analysis techniques to use. Here we provide a brief description of statistical techniques suitable for simple randomisation studies.

Parametric methods

For a study using a parallel groups design and an interval scale response variable (e.g. weight or blood pressure), the independent samples t-test will be used to compare two groups and one-way analysis of variance to compare three or more groups. For the two-period cross-over study, a refinement of the paired t-test is available, suggested by Hills and Armitage (2004), which takes account of the variability attributable to period effects and provides a test for carryover. If baseline values of a response variable are available, then changes in the variable during the intervention may be calculated and used in the analysis. However, if the baseline response values are not highly correlated with the final response values, then it can be more beneficial to analyse the final value in an analysis of covariance with the initial value considered as the covariate. For studies that take more than two serial measurements of response variables, the derivation of a summary measure such as a slope or area under the curve may permit the application of straightforward statistical techniques and avoid the need for more complex methods for correlated responses. Intervention effects, often expressed as means or differences in means, should be estimated along with their associated 95% confidence intervals.

Non-parametric methods

For ordinal scale outcomes non-parametric methods are typically employed, with the Mann–Whitney U test used to compare two groups, and Kruskal–Wallis one-way analysis of variance of ranks to compare three or more groups. However, these techniques focus on hypothesis testing, and confidence limits associated with them are not widely available. Non-parametric methods may also be useful for analysing interval scale variables for which the assumptions necessary for parametric methods are in doubt. Particularly in small studies, the assumption of normality in the distribution of the response variable is important. However, in such situations it may be possible to avoid resorting to non-parametric methods by transforming the data (often using a logarithmic transformation to reduce the degree of positive skew) prior to applying a parametric method.

Contingency table methods

For nominal scale (or unordered categorical) outcome variables, analysis is performed using chi-squared tests for contingency tables or Fisher's exact probability test where numbers are small. Confidence intervals for proportions, for differences in proportions, for odds ratios or for risk ratios may also be useful for characterising intervention effects.

If information on covariates is available, then it may be incorporated into an analysis of covariance to improve the precision of comparisons between intervention groups for an interval scale response. The technique does assume that there is a linear relationship between the response and the covariate in each group and that the linear relationships are parallel in the groups, assumptions that should be checked prior to using the method. It may also be useful in adjusting for chance imbalances between the intervention groups on factors relevant to the response. For a two-category response variable, logistic regression analysis may be employed in a similar way.

The interpretation of analyses involving more than two intervention groups may be complicated by the multiplicity of statistical tests. If the aim of an analysis is restricted to making only a small number of pre-specified comparisons between groups, as stated in the study protocol, then multiple testing is less of an issue. However, tests of hypotheses other than these (e.g. hypotheses formulated after looking at the results) require a more conservative approach in the statistical analysis to limit the risk of false positive findings. A similar issue arises in the interpretation of tests on multiple response variables. Ideally investigators should nominate the primary outcome measure in the study protocol. Other responses may still be analysed, but a stricter significance level may be appropriate to safeguard against false positive findings.

A recent development in nutrition research has been to use genomics, proteomics and metabolomics approaches as endpoints in nutrition intervention studies (see Chapter 13). Such studies often have multiple endpoints and no prior hypotheses, which raises similar statistical issues. If the multiple endpoints are independent, then a simple Bonferroni correction is sufficient to control the risk of type 1 error, with a significance level set not at the α level but at the α/k level, where k is the number of endpoints. An alternative approach, which retains more power than the Bonferroni correction and is more suited to microarray work, is to control the false discovery rate; that is, the expected proportion of false positives among the results that are declared significant. For dependent endpoints, comparisons are better performed by a permutation test. This involves comparing the largest test statistic obtained in the analyses of the various endpoints, not with a standard distribution (such as the t distribution or chi-squared distribution), but instead with its permutation distribution, obtained by calculating the largest test statistic in every possible random relabelling of the groups (or at least in a very large random sample of them).

Intention to treat or per protocol

An important issue in the analysis of interventions is to decide how protocol deviations should be handled. Usually the most relevant comparison of interventions will include all randomised participants who began the intervention, and the analysis will be conducted on an 'intention to treat' (ITT) principle. In an ITT analysis, once participants have been randomised to intervention groups, all available results are analysed in the groups to which they were allocated, regardless of whether or not the participants complied with the intervention. In nutrition studies there is often interest in examining response in the subset of participants who showed the best, or different levels of, compliance with the intervention (for a discussion of adequate levels of compliance see earlier in this chapter) and a 'per protocol' (PP) analysis may then be more relevant, even though this approach has a greater potential for introducing bias into the comparison of interventions.

Interpretation

The interpretation of study findings, and the discussion section of a resulting publication, should include a consideration of the study limitations, including any potential sources of bias (for example imbalance in baseline characteristics), imprecision (in outcome assessments) or an acknowledgement of the possibility of spurious statistically significant findings arising from multiple comparisons. The generalisability of the study findings should also be considered and limitations acknowledged. Conclusions should be confirmed and justified by the accompanying data. The conclusions should relate directly to the hypothesis, to the intervention at the dose or amount consumed, and to the population included in the study. Conclusions about secondary outcome measures should be stated as such and interpreted appropriately.

Roles and responsibilities of the research team

Complex issues arise because of potential conflicts of interest and scientific bias, particularly when research funding may come from the food industry. Many journals now require statements of the roles and responsibilities of all members of the research team, including the funders or sponsors, and declarations of any potential conflicts of interest. This should be standard practice when publishing any intervention study.

3.4 Conclusion

Intervention studies are a vital part of nutrition research, as if well designed they allow the testing of causality. Nutrition intervention studies vary considerably in study

design and duration, but there are a number of key design factors that must be considered when planning such a study, including the research question or hypothesis; duration; the intervention nutrient, food or diet; the intervention dose or amount; the control arm and blinding of the control; the primary and secondary outcome measures (including assessment of background diet); eligibility criteria; data-collection methodology; and measuring and encouraging compliance. Early involvement of a statistician in the study team to guide on both study design and statistical analysis is crucial. Local ethical approval and research governance procedures must be followed, and intervention studies registered before recruitment starts on a publicly accessible database. Finally, when reporting the results of the intervention, interpretation should be appropriate and any potential conflicts of interest, for example when funding has come from the food industry, should be declared.

Acknowledgements

JVW, RWW and CCP were members of the Expert Group on Guidelines for Human Intervention Studies to Scientifically Substantiate Claims on Foods, which was a working group of the ILSI Europe Functional Foods Task Force, and acknowledge the discussions of that group, which led to a scientific publication (listed in the references) and was the basis of the guidance presented here.

References and further reading

Archuleta, M., Vanleeuwen, D., Halderson, K. *et al*. (2012) Cooking schools improve nutrient intake patterns of people with type 2 diabetes. *Journal of Nutrition Education and Behavior*, **44** (4), 319–325.

Blundell, J., de Graaf, C., Hulshof, T. *et al*. (2010) Appetite control: Methodological aspects of the evaluation of foods. *Obesity Reviews*, **11** (3), 251–270.

Bonell, C.P., Hargreaves, J., Cousens, S. *et al*. (2011) Alternatives to randomisation in the evaluation of public health interventions: Design challenges and solutions. *Journal of Epidemiology and Community Health*, **65**, 582–587.

Brink, E., Coxam, V., Robins, S. *et al*. (2008) Long-term consumption of isoflavone-enriched foods does not affect bone mineral density, bone metabolism, or hormonal status in early postmenopausal women: A randomized, double-blind, placebo controlled study. *American Journal of Clinical Nutrition*, **87**, 761–770.

Brouns, F., Bjorck, I., Frayn, K.N. *et al*. (2005) Glycaemic index methodology. *Nutrition Research Reviews*, **18** (1), 145–171.

Carter, M.C., Burley, V.J., Nykjaer, C. and Cade, J.E. (2013) Adherence to a smartphone application for weight loss compared to website and paper diary: Pilot randomised controlled trial. *Journal of Medical Internet Research*, **15** (4), e32.

Craig, P., Dieppe, P., Macintyre, S. *et al*. (2008) Developing and evaluating complex interventions: The new Medical Research Council guidance. *British Medical Journal*, **337**, a1655.

Cummins, S., Petticrew, M., Higgins, C. *et al*. (2005) Large scale food retailing as an intervention for diet and health: Quasi-experimental evaluation of a natural experiment. *Journal of Epidemiology and Community Health*, **59**, 1035–1040.

Davis, P.A., Vasu, V.T., Gohil, K. *et al*. (2007) The effects of high walnut and cashew nut diets on the antioxidant status of subjects with metabolic syndrome. *European Journal of Nutrition*, **46**, 155–164.

Dekker Nitert M., Barrett1, H.L., Foxcroft, K. *et al*. (2013) SPRING: An RCT study of probiotics in the prevention of gestational diabetes mellitus in overweight and obese women. *BMC Pregnancy and Childbirth*, **13**, 50.

Erdmann, J., Hebeisen, Y., Lippl, F. et al. (2007) Food intake and plasma ghrelin response during potato-, rice- and pasta-rich meals. *European Journal of Nutrition*, **46**, 196–203.

Estruch, R., Ros, E., Salas-Salvadó, J. et al. (2013) Primary prevention of cardiovascular disease with a Mediterranean diet. *New England Journal of Medicine*, **368**, 1279–1290.

European Medicines Agency (1995) *Note for Guidance on Clinical Safety Data Management: Definitions and Standards for Expedited Reporting* (CPMP/ICH/377/95), http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002749.pdf (accessed October 2013).

Friedman, L.M., Furberg, C.D. and DeMets, D.L. (2010) *Fundamentals of Clinical Trials*, 4th edn. Springer, New York.

Guess, B.W., Scholz, M.C. and Strum, S.B. (2006) Modified citrus pectin (MCP) increases the prostate-specific antigen doubling time in men with prostate cancer: A phase II pilot study. *Prostate Cancer and Prostatic Diseases*, **6**, 301–304.

Hennekens, C.H., Buring, J.E., Manson, J.E. *et al*. (1996) Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *New England Journal of Medicine*, **334**, 1145–1149.

Hills, M. and Armitage, P. (2004) The two-period cross-over clinical trial. *British Journal of Clinical Pharmacology*, **58**, S703–S716.

International Committee of Medical Journal Editors (n.d.) *Clinical Trial Registration*. http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html (accessed June 2014).

Kelly, M.T., Wallace, J.M.W., Robson, P.J. *et al*. (2009) Increased portion size leads to a sustained increase in energy intake over 4 d in normal-weight and overweight men and women. *British Journal of Nutrition*, **102**, 470–477.

Lee, Y.M., Skurk, T., Hennig, M. and Hauner, H. (2007) Effect of a milk drink supplemented with whey peptides on blood pressure in patients with mild hypertension. *European Journal of Nutrition*, **46**, 21–27.

Machin, D. and Fayers, P. (2010) *Randomized Clinical Trials: Design, Practice and Reporting*. Wiley-Blackwell, Chichester.

Matthews, J.N., Altman, D.G., Campbell, M.J. *et al*. (1990) Analysis of serial measurements in medical research. *British Medical Journal*, **300**, 230–235.

Moher, D., Hopewell, S., Schulz, K.F. *et al*. (2010) CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, **340**, c869.

Neukam, K., Stahl, W., Tronnier, H. *et al*. (2007) Consumption of flavanol-rich cocoa acutely increases microcirculation in human skin. *European Journal of Nutrition*, **46**, 53–56.

Peace, K.E. and Chen, D. (2011) *Clinical Trial Methodology*. Chapman & Hall/CRC, Boca Raton, FL.

Piantadosi, S. (2005) *Clinical Trials: A Methodological Perspective*, 2nd edn. John Wiley & Sons, Inc, Hoboken, NJ.

Pocock, S.J. (1983) *Clinical Trials: A Practical Approach*. John Wiley & Sons Ltd, Chichester.

Price, R.K., Keaveney, E.M., Hamill, L.L. *et al.* (2010) Consumption of wheat aleurone-rich foods increases fasting plasma betaine and modestly decreases fasting homocysteine and LDL-cholesterol in adults. *Journal of Nutrition*, **140**, 2153–2157.

Schulz, K.F., Altman, D.G. and Moher, D. (2010) CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, **340**, c332.

Welch, R.W., Antoine, J.-M., Berta, J.-L. *et al.* (2011) Guidelines for the design, conduct and reporting of human intervention studies to evaluate the health benefits of foods. *British Journal of Nutrition*, **106**, S2–S15.

Woodside, J.V., Campbell, M.J., Denholm, E.E. *et al.* (2006) Short-term phytoestrogen supplementation alters insulin-like growth factor profile but not lipid or antioxidant status. Journal of Nutritional Biochemistry, **17**, 211–215.

Woodside, J.V., Koletzko, B.V., Patterson, C.C. and Welch, R.W. (2013) Scientific standards for human intervention trials evaluating the health benefits of foods and their application to infants, children and adolescents. *World Review of Nutrition and Dietetics*, **108**, 18–31.

World Medical Association (2008) *Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects*, http://www.wma.net/en/30publications/10policies/b3/index.html (accessed October 2013).

Yancy, W.S. Jr, Foy, M., Chalecki, A.M. *et al.* (2005) A low-carbohydrate, ketogenic diet to treat type 2 diabetes. *Nutrition and Metabolism*, **2**, 34.