Guidelines for the Design, Conduct and Reporting of Human Intervention Studies to Evaluate the Health Benefits of Foods

Robert W. Welch[1], Jean-Michel Antoine[2], Jean-Louis Berta[3], Achim Bub[4], Jan de Vries[5], Francisco Guarner[6], Oliver Hasselwander[7], Henk Hendriks[8], Martin Jäkel[9], Berthold V. Koletzko[10], Chris C. Patterson[11], Myriam Richelle[12], Maria Skarp[13], Stephan Theis[14], Stéphane Vidry[13] and Jayne V. Woodside[11]

1. Northern Ireland Centre for Food and Health, School of Biomedical Sciences, University of Ulster, Coleraine BT52 1SA, UK
2. Danone Research, Route Départementale 128, FR-91767, Palaiseau Cedex, France
3. Consultant, 37 rue Boileau, FR-92120 Montrouge, France
4. Max Rubner-Institute, Federal Research Centre for Nutrition and Food, Haid-und-Neu-Strasse 9, DE-76131, Karlsruhe, Germany
5. De Vries Nutrition Solutions, Reuvekamp 26, 7213 CE Gorssel, The Netherlands
6. Digestive System Research Unit, Hospital General Vall d'Hebron, CIBEREHD, Passeig Vall d'Hebron, 119-129, ES-08035 Barcelona, Spain
7. Danisco, 43 London Road, Reigate, RH2 9PW, Surrey, UK
8. TNO Quality of Life, PO Box 360, NL-3700 AJ Zeist, The Netherlands
9. Unilever, Olivier van Noortlaan 120, NL-3133 AT Vlaardingen, The Netherlands
10. Division of Metabolic and Nutritional Medicine, Dr von Hauner Children's Hospital, University of Munich Medical Centre, Lindwurmstrasse, 4, DE-80337 Munich, Germany
11. Centre for Public Health, Queen's University Belfast, Belfast BT12 6BJ, UK
12. Nestlé Research Center, PO Box 44, Vers-chez-les-blanc, CH-1000 Lausanne, Switzerland
13. ILSI Europe a.i.s.b.l., Avenue E. Mounier 83, Box 6, BE-1200 Brussels, Belgium
14. Südzucker/BENEO Group, Wormserstrasse 11, DE-67283 Obrigheim, Germany

ILSI Europe
International Life Sciences Institute

Commissioned by the
ILSI Europe Functional Foods Task Force

Correspondence: ILSI Europe a.i.s.b.l. - Avenue E. Mounier 83, Box 6 - B-1200 Brussels - Belgium
email publications@ilsieurope.be - Fax: +32 2 762 00 44

Vol. 106   Supplement No. S2   November 2011   **British Journal of Nutrition**

Table of Contents

Correspondence: ILSI Europe a.i.s.b.l., Avenue E. Mounier 83, Box 6 - 1200 Brussels, Belgium, fax: +32 2 762 00 44, email: publications@ilsieurope.be

Abbreviations: AE, adverse event; CONSORT, Consolidated Standards of Reporting Trials.

There is substantial evidence to link what we eat to the reduction of the risk of major chronic diseases and/or the improvement of functions. Thus, it is important for public health agencies and the food industry to facilitate the consumption of foods with particular health benefits by providing consumer products and messages based on scientific evidence. Although fragmentary advice is available from a range of sources, there is a lack of comprehensive scientific guidelines for the design, conduct and reporting of human intervention studies to evaluate the health benefits of foods. Such guidelines are needed both to support nutrition science in general, and to facilitate the substantiation of health claims. In the present study, which presents the consensus view of an International Life Sciences Institute Europe Expert Group that included senior scientists from academia and industry, the term 'foods' refers to foods, dietary supplements and food constituents, but not to whole diets. The present study is based on an initial survey of published papers, which identified the range and strengths and weaknesses of current methodologies, and was finalised following exchanges between representatives from industry, academia and regulatory bodies. The major factors involved in the design, conduct and reporting of studies are identified, summarised in a checklist table that is based on the Consolidated Standards of Reporting Trials guidelines, and elaborated and discussed in the text.

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[1–3], as well as the improvement of functions, for example muscle function or immune response[4,5]. Thus, it is important for public health agencies and the food industry to be aware of these links and to provide messages and products that will facilitate the consumption of healthy diets by consumers. However, it is important that these messages and products are supported by good scientific evidence.

The aim of the present study is to provide guidelines for the design, conduct and reporting of human intervention studies. These guidelines should assist with studies designed to support nutrition science in a broad sense, and also aim to substantiate health claims for foods. In the present study, the term 'foods' is used to mean foods, dietary supplements and food constituents, but does not cover whole diets.

These guidelines, which are the consensus view of an International Life Sciences Institute Europe Expert Group, were finalised following exchanges between representatives from industry, academia and regulatory bodies[6]. The Expert Group carried out an initial survey of relevant research papers published within pre-defined periods in two leading peer-reviewed journals (*Am J Clin Nutr*; *Eur J Nutr*). This database was subsequently augmented with selected research papers, to provide examples not present in the initial selection. This survey facilitated the identification of the range, and the perceived strengths and weaknesses of currently reported methodologies, and these guidelines cite relevant selected papers as examples of current practice.

The major factors involved in the design, conduct and reporting of studies are identified in Table 1, which uses a similar structure to that in the Consolidated Standards of Reporting Trials (CONSORT) checklist for medical trials[7]. These factors are discussed in the text under the same main headings used in Table 1.

Summary of existing guidelines

Although earlier guidelines exist[8], newer guidelines are emerging in response to legal requirements[9]. These guidelines include those aimed at particular products and health benefits[10–12]. However, there is a lack of a comprehensive description on how to perform human nutrition intervention studies in general, and particularly for the evaluation and substantiation of health claims on foods. Besides the 'PASSCLAIM (Process for the Assessment of Scientific Support for Claims on Foods) Consensus Criteria'[13], which so far is the most conclusive summary, various collections of advice exist, which are given fragmentarily in legal regulations or guidance reports from international and national authorities or organisations. These are outlined below.

European Food Safety Authority Scientific and Technical Guidance[14] and the 'Application Rules on Health Claims' in commission Regulation EC 353/2008[15] provide only minor information on the conduct of human intervention studies. Within Europe, additional information is available from the Joint Health Claims Initiative[16], where some criteria on the validity of human studies are listed. The 'US Food and Drug Administration Guidance for Industry' (2009)[17] gives a broad description on human intervention studies in a question and answer style, which guides through relevant human intervention issues. Important aspects for claim evaluation can also be drawn from the updated 'Health Canada Guidance Document' (2009)[18]. The varying approaches to health claim evaluation in Europe, the USA, Canada, Australia and New Zealand, China and Japan have quite recently been reviewed in a supplement issue of *J Nutr*[19]. Additionally, the 'FAO/WHO Codex Alimentarius Commission' (2008)[20] gives basic criteria for health claim evaluation. However, a concise summary of evaluation criteria focusing on the conduct of human intervention studies is not currently available. An attempt in this direction may be the 'Nutrition, Health and Related Claims Consultation Paper (Proposal P293, to be finalized by late 2011)'[21] by 'Food Standards Australia New Zealand', which may provide details on the evaluation and substantiation requirements for health claims.

Factors to be considered in the design, conduct and reporting of human intervention studies

Study hypothesis

The primary hypothesis to be tested will directly influence all other aspects of the study, including the study design and

Table 1. Checklist: factors to be considered, and recommendations for best practice when designing, conducting and reporting human intervention studies to evaluate the health benefits of foods

Phase	Factors to consider	Recommendations for design and conduct	Recommendations for reporting
Design	Hypothesis	Clear hypothesis	Explicitly state hypothesis, link to primary outcome measures
	Study design	Appropriate design	Clearly describe, with rationale
	Duration	Appropriate to design, intervention and outcome measures	Clearly describe, with rationale
	Intervention	Test and control products suitably matched	Describe test and control products in detail, with rationale
	Amount	Appropriate to outcome measures and to practical usage	Clearly describe, with rationale
	Outcome assessment	Define primary outcomes and methods of measurement	Clearly describe how and when assessed and link to hypothesis
		Define all secondary outcomes and methods of measurement	Clearly describe how and when assessed
	Eligibility criteria	Define all eligibility criteria	Describe criteria using objective, quantitative descriptors where possible
	Statistical considerations		
	Randomisation	Use randomised design where possible and ensure appropriate method for allocation sequence generation and concealment	Clearly describe randomised design and the methods used for randomisation, sequence generation and concealment
	Blinding	Ensure double blinding if feasible, single blinding if not	Describe how blinding was achieved (who was blinded and how), report success rate
	Size of study	Conduct power calculation based on primary outcome measures	Include all elements of power calculation
Conduct	Study protocol		
	Ethical approval and trial registration	Obtain full ethical approval, register trial, comply with the Declaration of Helsinki	Give details of research ethics authority and approval number, and database and registration number
	Recruitment	Define recruitment strategy and process, including settings and dates	Explicitly describe strategy, provide participant flow diagram
	Data collection		
	Background diet and monitoring change	Select suitable methods to collect and analyse data	Describe assessment and analysis methods, report descriptive data on background diet and changes for all components that may be relevant by allocated intervention group
	Background health status and lifestyle, and monitoring changes	Define relevant measures, select suitable methods of assessment	Justify relevant measures, describe assessment methods, and report relevant factors and changes by allocated intervention group
	Unintended effects	Devise strategy and methods to capture data	Report methods to assess unintended effects and report by allocated intervention group
	Adverse events	Have mechanisms in place to record and respond to adverse events	Clearly define and report all adverse events by allocated intervention group
	Compliance	Define acceptable levels of compliance, use appropriate strategies to maximise compliance, select and use rigorous but feasible methods for assessment of compliance	Report methods used to measure and maximise compliance, report compliance rates numerically and by allocated intervention groups
Analysis and interpretation	Statistical analysis	Devise appropriate analysis methods, based on study design and outcome measures	Describe distribution of data, present descriptive characteristics by allocated intervention group, present hypothesis tests for comparing allocated intervention groups, make clear distinction between primary v. secondary endpoint analyses, state whether analysis ITT or PP
	Discussion and interpretation	Consider study limitations and generalisability of findings	Discussion of limitations and generalisability of study findings
	Conclusion	Relate directly to hypothesis, study design, test product and study participants	Clear statement of conclusion

ITT, intention to treat; PP, per protocol.

duration, the eligibility criteria, the amount of the test product and the nature of the control. The hypothesis should be based on a thorough review of the available evidence. The primary outcome measure should be clearly defined and relate to the hypothesis.

Current practice. The review of current practice showed that only a minority of the papers mentioned the term 'hypothesis', with the majority framing the research question as 'aims', 'objectives' or 'purposes'. While hypotheses can sometimes be inferred from aims or objectives, this is not always the case and therefore it is recommended that studies have a clearly stated hypothesis.

Study design

Human nutrition intervention studies test hypotheses that have been formulated based on prior knowledge. Prior knowledge will include data from other human intervention studies, and also epidemiological, animal and *in vitro* studies. Where possible, all the available evidence should be reviewed systematically[22] for efficacy, and include an assessment of safety and potential risks.

Exploratory studies generally have a number of aims. These may include evaluations of food matrix issues and the amount to be consumed, and can also provide data on the variability and time scale of outcome responses and the size of the effect on outcomes responses, which can be used for power calculations in subsequent studies.

Once these early studies have been completed, studies with greater rigour will test the hypothesis that the product will alter the expected 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 to be consumed and evaluating different forms of presentation or delivery of the test product.

The following three basic study designs are encountered: single-arm studies; cross-over studies; parallel studies. Early exploratory studies tend to be single arm with no control group, and these can be a cost- and time-effective way of assessing potential effects, but usually only as a forerunner to controlled studies. These studies add to the totality of evidence but cannot alone determine the effect of intervention.

In controlled studies, in addition to a group of participants receiving the active nutrition intervention, there will be another group that will act as a control. The outcome for this latter group provides a suitable comparator, as it is generally inappropriate to assume that changes observed in the group receiving the active intervention during the study are necessarily entirely attributable to that particular intervention; other factors may be responsible for them. For example, the knowledge that a participant is receiving an active intervention may alter their responses, particularly when dealing with more subjective outcomes such as quality of life scales.

In parallel-group designs, each participant receives only one of the nutrition interventions (product A or B, active intervention or control) 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 studies using cross-over designs, 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 precision of the comparisons and therefore power of the study. In such designs, participants act as their own controls. In a cross-over design for two interventions, the participants are allocated to two groups which 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. Also a run-in period may be desirable in advance 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, any increase in study length may increase participant dropout 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 may be affected. 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 satiety and 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 the availability of time and other resources, and the potential roles of confounding factors such as seasonal variations.

Other less commonly used types of study include (1) the factorial design (in which participants are allocated to all possible combinations of two or more interventions and which permits the evaluation of intervention interactions) and (2) 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[23–27].

Study duration

The study duration should be sufficient to allow changes in the primary outcome measure. Thus, the duration will be informed by data from exploratory studies, from knowledge of the underlying physiology and biochemistry, particularly

the rates of turnover of relevant tissues such as erythrocytes, or from similar studies that have used the same outcomes. It also relates to the envisaged claim or hypothesis, which may focus on an acute effect (e.g. glycaemic response or increased alertness) *v.* a longer-term health outcome. Thus, no absolute guidelines can be given on study duration. However, researchers should aim to use the minimum 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.

Test and control product

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

Amount consumed.   The amount of the test product to be consumed will depend on a number of factors. These include evaluation of data from all previous studies, and consideration of the underlying physiology and of issues related to the 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 give relevant documentary evidence of the amount of the test product or the component with putative activity that is provided.

Control product.   The control is a product that does not provide the component that is being tested, and this must also be analytically documented. The control should be matched for sensory characteristics and taken in the same way as the test product. A control is relatively easy to achieve in supplementation studies using pills or similar preparations, but in studies of foods, it is more difficult to develop a control product identical to the test product but which does not contain the component(s) under study. Blinding may not be possible for many foods where the test product is easily identifiable by the researchers, as may be the case with some minimally processed foods such as fruits or vegetables, and some manufactured consumer foods such as cereal products. However, some degree of blinding may be possible by the use of suitable packaging that conceals products from the researchers and other study participants.

Success in attaining an ideal control is likely to vary depending on the type of ingredient or product under test. Ideal controls for different product types are considered below.

Supplements (pills, powders, liquids of small volume). The provision of high-quality control products should be relatively easy. Pills should be of identical size, shape, colour and appearance. Ensuring identical internal colour, appearance, mouthfeel and taste for all supplements is also desirable.

Food ingredients (e.g. fibres, starches, proteins, fats). These can be evaluated in the form of supplements (see the previous section) or incorporated into suitable consumer foods (see the next section). The potential effects of interaction with other food components, and/or potential effects of processing (e.g. degree of loss or modification, effects on bioavailability) must be considered.

Manufactured consumer foods (e.g. cereal products, juices, prepared dishes, yogurts).   Test and control foods should be similar in energy content and in physical characteristics (gross morphology, appearance, volume and texture) and sensory qualities (mouthfeel, taste, palatability and breakdown characteristics in the mouth). If identical products cannot be produced, then it is possible that compositional, physical or sensory differences in the test and control products, unrelated to the factor under test, may exert confounding physiological or behavioural effects unrelated to the factor under test, and will also make effective blinding difficult. In these circumstances, appropriate physiological or behavioural responses should be monitored and compared for the test and control products.

Minimally processed foods (e.g. fresh fruit, vegetables, nuts, eggs, grains).   The formulation of control products is impossible if the aim is to use a single foodstuff, such as fruit or nuts. If the control arm only involves the provision of no test product or a smaller number of portions of the test product, then this may have effects on other aspects of diet and behaviour.

Outcome measures

All intervention studies will assess outcome measures, and will compare these between intervention test product 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 a pre-specified primary outcome and the sample size calculated based on that outcome (see the Size of study section). 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 measure. All outcomes 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 intended effect, for example body weight, or diagnosis of a disease or muscle strength. Subjective measures are also used, such as feelings of health, appetite or fatigue, and, 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 biomarkers or risk factors are used that reflect a functional, physiological or biochemical characteristic associated with a function or a disease, or that predict later development of the disease. Examples include glycated Hb as an indicator of long-term hyperglycaemia and risk of type 2 diabetes complications[28], plasma LDL-cholesterol as a measure of CVD risk[29], bone mineral density as a measure of osteoporosis risk[30], complex metabolomic or proteomic profiles as markers of function and disease risk[31], and the

presence of adenomatous colon polyps as an early indicator of colon cancer[32]. Most indirect outcome measures are chosen because they reflect consensus guidelines or are commonly used by experts in the area. However, very few markers have been validated and recognised by expert consensus in terms of their specificity, variability, limitations and applicability to a range of population groups[13].

Methodological aspects. Efforts 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.

Analytical variability. Laboratory analytical methods should be precise, accurate, sensitive and specific, and these performance characteristics should be recorded in standard operating procedures or similar quality record documents. 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. Inter-laboratory analytical variability, which may be a factor in multi-centre studies, should either be minimised by the sharing of methods, standard operating procedures and calibration standards, or be overcome by centralisation of sample analysis. Where appropriate, statistical analysis should be used to account for any remaining inter-laboratory variation. Biomarkers that have high methodological variability will often require a higher number of participants.

Biological variability. Biological variability has a number of underlying factors, and such variability is likely, in part, to be genetically determined. Many biomarkers, similar to most biological parameters, have circadian or seasonal variations. The basal value can also fluctuate due to biological rhythms, such as the menstrual cycle. This variability may introduce systematic bias into the results. 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. A study may show a statistically significant change in a validated biomarker. However, a statistically significant response of an outcome measure to a nutrition intervention does not necessarily mean that the intervention will be effective in terms of benefit or risk reduction in target groups. Therefore, the size of changes and whether these will be of biological, clinical or public health significance must also be considered.

Selection of participants: eligibility criteria

Eligibility criteria 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 intakes, or the consumption of restricted diets. Eligibility criteria can be presented as inclusion and exclusion criteria. The criteria are likely to include factors such as age, sex, health status, and underlying physiological conditions or concurrent diseases.

Eligibility criteria should describe participants adequately, so that the results can be appropriately interpreted in terms of their generalisability. Eligibility criteria should be selected with the target population for the test product, and the hypothesis and outcome measures in mind. Inter-participant variation may usefully be reduced by using stricter eligibility criteria to select a more homogeneous set of participants for study, but this also has the disadvantage of restricting the target population and consequently limiting the generalisability of findings. Children and women of childbearing age may need to be excluded from studies of certain interventions with developmental implications or 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[33]. 'Health' may just refer to the absence of diagnosed disease, or refer to a specific aspect such as a healthy blood pressure, and in such cases, the criteria can be very specific[34] 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.

Current practice. The majority of human intervention studies report and define eligibility criteria to some extent. However, most papers do not give all eligibility criteria, quantifiable ranges for these criteria or a clear rationale for these criteria, and do not relate the criteria to the hypothesis being tested. In addition, many studies mention 'healthy' as an inclusion criterion without defining 'health' status. An example of good practice is Brink *et al.*[35], where inclusion and exclusion criteria are described with quantifiable ranges for many of the criteria.

Statistical considerations

Randomisation. Randomisation is the allocation of interventions to participants using some random process such as the toss of a coin. Randomisation ensures that the investigator does not influence the intervention to which a participant is allocated. The main advantage of random allocation is that, in the long run, it will produce study groups, which are comparable with respect to both known and unknown factors, which could influence the outcome measure. Consequently, 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 (unbiased) and the statistical analysis is valid.

To allocate individual participants to intervention groups, random numbers (either from tables or generated by

computer) can be used. However, it is advisable to ensure that approximately equal numbers of participants are assigned to each group by using a block randomisation, in which participants are divided into blocks within which equal numbers of allocations to each intervention are made. 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, sex 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 in the statistical analysis. Minimisation offers a more practicable approach to stratification on multiple variables.

Concealment of the intervention allocations.   The decision to enrol a participant in a study could be influenced subconsciously or otherwise by the knowledge of which intervention the participant would receive if entered into the study. A simple way to eliminate any possible bias of this sort is to implement randomisation using sealed envelopes. The list of random intervention allocations is concealed in sequentially numbered opaque sealed envelopes. Only after a participant has been enrolled in the study and consent obtained should the seal be broken to reveal which intervention the participant has been allocated.

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 the assessor nor the 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, usually after the freezing of the database, should be documented in the study report and may be mentioned in the publication.

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. Currently, this information is rarely reported.

Size of study.   Estimation of the number of participants required for the study is essential. Too small a study is likely to fail to detect important differences between interventions, while too large a study may needlessly waste resources and be unethical. In certain circumstances, trials may be designed for interim analysis as each participant's result becomes available (sequential design) or after pre-specified numbers of participants' results become available (group-sequential designs)[23–27]. 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. 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 dropout rate. Information about the degree of variability in the outcome is also required and may come from previously published or unpublished results, or from a pilot or exploratory study specifically performed for the purpose. A multi-centre study may be necessary if the study size 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 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 Declaration of Helsinki[36]. One of the recommendations of the Declaration of Helsinki 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 (e.g. repetition of negative studies), selective reporting of outcomes, and of per-protocol rather than intention-to-treat analyses (see the Statistical analysis section). The WHO[37] has stated that 'the registration of all interventional trials is a scientific, ethical and moral responsibility' while the International Committee of Medical Journal Editors have, from September 2004, only considered trials for publication if they were registered before enrolment of their first participant[38]. The academic view is that *a priori* trial registration is essential for ethical research in human participants. For industry-based nutrition studies, where the results may be commercially sensitive or where intellectual property rights may be jeopardised, a position similar to that in the pharmaceutical industry could be adopted, whereby five items of protocol information are kept hidden in a locked electronic depository that is publicly inaccessible until the information is no longer deemed commercially sensitive[39].

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 participants approached, screened, recruited and completing, and reasons for non-recruitment (ineligibility, lack of willingness to participate) and non-completion noted. Informed consent should be obtained. This information is best summarised in a participant flow diagram when reporting the study[7].

Data collection

Data should be collected using standardised case report forms. Participants should be assigned a unique study number at the start of the study, and data held under that study number, i.e. there should be no participant identifiable information 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 regulations.

Background diet and change in diet during intervention. The nature of the participants' background diet may be one of the eligibility criteria. However, it is important, regardless of eligibility criteria, and particularly in longer-term studies, to collect background dietary information in order to characterise the habitual diet of the participants in terms of foods and diet composition. Diet should also be assessed during longer-term interventions, in order to detect any potentially confounding changes that may occur. A number of methodologies are available for these dietary assessments (e.g. FFQ, food diary, diet history[40]). However, dietary intake assessment is subject to misreporting, and reported energy intakes should be compared with the estimated energy requirements of participants, particularly if these assessments are used for monitoring compliance.

Background health status and lifestyle, and changes in health status and lifestyle during intervention. In addition to their possible roles 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 suitable interpretation and generalisation of the results. Examples include age, sex, level of medication use, 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. This is particularly important in studies where it is not possible to use closely matched test and control products.

Unintended effects. Aside from formal adverse events (AE), which are discussed below, in nutrition studies, there may be other unintended effects (to use recent CONSORT[7] terminology). Unintended effects arising during nutrition studies are likely to be restricted to, for example, mild nausea or minor gastrointestinal discomfort as a result of changes in dietary pattern or consumption of unfamiliar or blinded products.

The recording of these unintended effects is desirable in human nutrition interventions, providing data on the tolerability of the product and enabling the compilation of a dossier for future reference. Questionnaires should be used to provide quantifiable data, using 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.

Adverse events. An AE is any untoward medical occurrence or undesirable clinical experience in a participant in a clinical trial, whether or not considered related to the intervention. Recording AE is of major importance in pharmaceutical studies, allowing a risk–benefit analysis. Hence, there is an abundance of guidelines for the management of AE 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)[41–43]. There are no guidelines for nutrition intervention studies, given that these studies involve testing foods, supplements or ingredients in participants who are usually apparently healthy. However, the formal recording of AE is required for good practice in nutrition research.

Any serious or unexpected AE that are encountered, whether or not they appear related to the intervention, should be reported immediately to the principal investigator, the research ethics committee, the sponsor and relevant regulatory bodies in order to ensure 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, and cause the reporting of false negative findings, and ultimately a lack of evidence to support a potentially beneficial effect. Poor compliance in a particular subgroup will also reduce the generalisability of results. 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 measure and maximise 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 under supervision and thus, compliance is not usually an issue. However, maintaining compliance throughout longer-term studies is very important. Informing participants that compliance will be measured by one of the methods below is likely, in itself, to improve adherence to the dietary intervention.

Complete provision of intervention, consumed under supervision. This will maximise compliance, but requires access to a metabolic suite or equivalent, and there will be resource implications. The suitability of this approach is likely to depend on the duration of the intervention.

Complete provision of intervention, with the return of unconsumed items. The return of unconsumed items is often used, but there can be no certainty that the participant has actually consumed all unreturned items.

Assessment of biomarkers in biological samples. Where possible, an independent and objective measure of compliance should be used, for example assessment of Se in serum or fatty acid composition of erythrocyte membranes[10].

R. W. Welch *et al.*

Using dietary records. Weighed food intakes, food diaries, food checklists and dietary recall methods, including diet histories, can be used. An inherent weakness of self-reported dietary intake data is the prevalence of socially desirable responding, where participants tend to under-report overall dietary intake and to over-report intakes of the intervention foods, and 'healthy' foods in general. Dietary records can usefully augment objective biological markers. If valid biomarkers of exposure do not exist, it is important to try to ensure that intakes are reliably reported.

A combination of the above. As no single method of assessing compliance is completely infallible, a combination of methods may be required.

In addition to the compliance assessment methods above, maintaining close, regular contact with participants is key to achieving good compliance, allowing any issues to be identified and dealt with at an early stage.

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 section on Statistical analysis for discussion of how compliance will affect statistical analysis. A decision on the statistical analysis approach will be partly influenced by whether studies are designed as tests of efficacy (biological effect) or effectiveness (potential to modify outcome in 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 as detailed above. *A priori* decisions should be made regarding the acceptable level of compliance for inclusion in a per-protocol analysis.

Current practice. An acceptable level of compliance is rarely reported quantitatively, with the extent of compliance usually reported qualitatively. However, as an example of good practice, Brink *et al.*[35] monitored compliance by a combination of specially designed forms, return of unconsumed foods, and assessment of both plasma and urine concentrations of isoflavones. These authors reported that the level of compliance was high (94%) and did not differ between intervention groups. Plasma isoflavone concentrations increased significantly in the active intervention groups.

Statistical analysis

There are a number of statistics books that cover the basics of randomised intervention trial methodology and analysis[23–27]. It is good practice to have a statistical plan that specifies the statistical methods to be used, the hypotheses to be tested for both primary and secondary outcomes (including whether one-sided or two-sided), and the significance level to be employed.

Rationale for using statistical methodology. In common with other research in medicine and the biological sciences, differences between groups which the investigator wishes to identify in a nutrition study are usually masked by several types of variation (inter- and intra-participant variation, measurement error, etc.), and strategies to minimise these have been outlined earlier.

These sources of variation 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 CI 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 hospital outpatient 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 non-response.

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 in a comparative study will be 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. The following provides a brief description of statistical techniques suitable for simple randomisation studies.

Parametric methods. For a study using a parallel-group 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 ANOVA to compare three or more groups[23,25]. For the two-period cross-over study, a refinement of the paired *t* test, which takes account of variability attributable to period effects, is typically employed[44]. 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, although it can be

more beneficial to analyse the final value of the response in an ANCOVA 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 more complex methods for correlated responses[45]. Intervention effects, often expressed as means, or differences in means, should be estimated along with their associated 95 % CI.

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 ANOVA of ranks to compare three or more groups[23,25]. These techniques may also be used for analysing interval-scale response variables, which do not satisfy the assumptions for the parametric methods. However, these techniques focus on hypothesis testing, and confidence limits associated with these techniques are not widely available.

Contingency table methods. For nominal-scale (or unordered categorical) outcome variables, the analysis is performed using χ^2 tests for contingency tables or Fisher's exact probability test where numbers are small. CI for proportions, for differences in proportions, for OR or for risk ratios may also be useful for characterising intervention effects.

If information on covariates is available, then it may be incorporated into a multiple regression analysis to improve the precision of comparisons between interventions on an interval-scale response. This technique may also be a useful approach in adjusting for chance imbalances between the intervention groups on factors relevant to the response. Similarly, for a categorical response variable, logistic regression analysis may be used.

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 response variable 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[31]. The use of multiple endpoints such as these raises some 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 a level but at the a/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, the expected proportion of false positives among the results that are declared significant. For dependent endpoints, comparisons may be performed by a permutation test. This involves comparing the largest test statistic, not with a standard distribution (such as the t distribution or χ^2 distribution), but instead with its permutation distribution obtained by calculating the largest test statistic in every possible 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 is to decide how protocol deviations should be handled in the analysis. 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' principle. 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, interest sometimes focuses on the subset of participants who showed good compliance with the intervention (for a discussion of adequate levels of compliance see the Compliance section) and a 'per-protocol' analysis may then be more relevant even though this approach has a greater potential for introducing bias into the comparison of interventions.

Discussion and interpretation

The interpretation of study findings, and discussion section of a paper should include a consideration of the study limitations, including any potential sources of bias (e.g. imbalance in baseline characteristics), imprecision (in outcome assessments) or an acknowledgement of the possibility of statistically significant findings arising from multiple comparisons. The generalisability of the study findings should also be considered and limitations acknowledged.

Conclusions

The conclusions should be confirmed and justified by the accompanying data. The conclusions should relate directly to the hypothesis, to the test product at the amount consumed and to the population included in the study. Conclusions about secondary outcome measures should be stated as such and interpreted appropriately.

Current practice. In practice, the hypothesis is often not clearly stated, and this can make the validity of the conclusions difficult to judge. Sometimes, statistically significant results are overemphasised when these were not the original focus of the study design. Another issue that occurs frequently is that findings of studies are generalised to broader populations than may be reasonable given the participant characteristics. Inappropriate generalisation of study findings can also occur in relation to the specific product and the amount consumed, and to the duration of the study.

Roles and responsibilities of the research team

The complex issues involved in potential conflicts of interest and scientific bias, particularly when research funding may come from the food industry, have recently been discussed[46]. Furthermore, 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. We recommend that this should be a standard practice.

Acknowledgements

References

1. Brunner EJ, Rees K, Ward K, *et al.* (2007) Dietary advice for reducing cardiovascular risk. *The Cochrane Database of Systematic Reviews*, issue 4, CD002128.

2. Van Horn L, McCoin M, Kris-Etherton PM, *et al.* (2008) The evidence for dietary prevention and treatment of cardiovascular disease. *J Am Diet Assoc* **108**, 287–331.

3. World Cancer Research Fund (2007) Food, Nutrition, Physical Activity, and the Prevention of Cancer. A Global Perspective. http://www.dietandcancerreport.org/ (accessed October 2010).

4. Asp N-G, Cummings JH, Mensink RP, *et al.* (2003) PASSCLAIM Phase one: preparing the way. *Eur J Nutr* **42**, Suppl. 1, 1–119.

5. Asp N-G, Cummings JH, Howlett J, *et al.* (2004) PASSCLAIM Phase two: moving forward. *Eur J Nutr* **43**, Suppl. 2, 1–183.

6. Aggett P, Antoine J-M, de Vries J, *et al.* (2010) Beyond PASSCLAIM – guidance to substantiate health claims on foods. Summary report of an ILSI Europe workshop held in December 2009 in Nice, France. http://www.ilsi.org/Europe/Publications/Beyond%20PASSCLAIM%20Report%202010.pdf (accessed August 2010).

7. Moher D, Hopewell S, Schulz KF, *et al.* (2010) CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* **340**, c869.

8. Sandstrom B (1995) Quality criteria in human experimental nutrition research. *Eur J Clin Nutr* **49**, 315–322.

9. European Parliament and Council (2007) Corrigendum to regulation (EC) no. 1924/2006 of the European Parliament and of the council of 20th December 2006 on nutrition and health claims made on foods. *OJ* **L12**, 3–18. http://eur-lex.europa.eu/LexUriServ/site/en/oj/2007/l_012/l_01220070118en00030018.pdf (accessed October 2010).

10. AbuMweis SS, Jew S & Jones PJ (2010) Optimizing clinical trial design for assessing the efficacy of functional foods. *Nutr Rev* **68**, 485–499.

11. Klein MA, Nahin RL, Messina MJ, *et al.* (2010) Guidance from an NIH workshop on designing, implementing, and reporting clinical studies of soy interventions. *J Nutr* **140**, 1192S–1204S.

12. Shane AL, Cabana MD & Vidry S (2010) Guide to designing, conducting, publishing and communicating results of clinical studies involving probiotic applications in human participants. *Gut Microbes* **1**, 243–253.

13. Aggett PJ, Antoine J-M, Asp N-G, *et al.* (2005) PASSCLAIM process for the assessment of scientific support for claims on foods: consensus on criteria. *Eur J Nutr* **44**, Suppl. 1, 1–30.

14. EFSA (2007) Scientific and technical guidance for the preparation and presentation of the application for authorisation of a health claim. *EFSA J* **530**, 1–44.

15. European Commission (2008) Commission regulation (EC) No 353/2008 establishing implementing rules for applications for authorisation of health claims as provided for in article 15 of regulation (EC) No 1924/2006 of the European Parliament and of the council. *OJ* **L109**, 11–16. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:109:0011:0016:EN:PDF (accessed October 2010).

16. Joint Health Claims Initiative (2000) Code of Practice on Claims on Foods. http://www.jhci.org.uk/info/code.pdf (accessed October 2010).

17. US FDA (2009) Guidance for Industry: Evidence-Based Review System for the Scientific Evaluation of Health Claims. http://www.fda.gov/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/FoodLabelingNutrition/ucm073332.htm (accessed October 2010).

18. Health Canada, Bureau of Nutritional Sciences (2009) Guidance Document for Preparing a Submission for Food Health Claims. http://www.hc-sc.gc.ca/fn-an/alt_formats/hpfb-dgpsa/pdf/legislation/health-claims_guidance-orientation_allegations-sante-eng.pdf (accessed October 2010).

19. Jones PJH, Asp N-G & Silva P (2008) Evidence for health claims on foods: how much is enough? Introduction and general remarks. *J Nutr* **138**, S1192–S1227.

20. Codex Alimentarius Commission (2008) Proposed Draft Annex to the Codex Guidelines for the Use of Nutrition and Health Claims: Recommendations on the Scientific Substantiation of Health Claims. CRD 17. November 2008. http://www.ccnfsdu.de/fileadmin/user_upload/PDF/2008/CRD_17.pdf (accessed October 2010).

21. Food Standards Australia New Zealand (2008) Nutrition, Health and Related Claims – Final Assessment Report – Proposal P293. http://www.foodstandards.gov.au/_srcfiles/P293 %20Health%20Claims%20FAR%20Attach%204 %20FINAL.pdf (accessed October 2010).

22. Moher D & Tricco AC (2008) Issues related to the conduct of systematic reviews: a focus on the nutrition field. *Am J Clin Nutr* **88**, 1191–1199.

23. Pocock SJ (1983) *Clinical Trials: A Practical Approach*. Chichester: John Wiley & Sons.

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

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

26. Friedman LM, Furberg CD & DeMets DL (2010) *Fundamentals of Clinical Trials*, 4th ed. New York: Springer.

27. Peace KE & Chen D (2011) *Clinical Trial Methodology*. Boca Raton, FL: Chapman & Hall/CRC.

28. UK Prospective Diabetes Study Group (1998) Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* **352**, 837–853.

29. Mensink RP, Aro A, Den Hond E, *et al.* (2003) PASSCLAIM – diet-related cardiovascular disease. *Eur J Nutr* **42**, Suppl. 1, 6–27.

30. Prentice A, Bonjour JP, Branca F, *et al.* (2003) PASSCLAIM – bone health and osteoporosis. *Eur J Nutr* **42**, Suppl. 1, 28–49.

31. Zhang X, Yap Y, Wei D, *et al.* (2008) Novel omics technologies in nutrition research. *Biotechnol Adv* **26**, 169–176.

32. Winawer SJ, Zauber A, Ho MN, *et al.* (1993) Prevention of colorectal cancer by colonoscopic polypectomy. *N Eng J Med* **329**, 1977–1981.

33. Perez-Martinez P, Yiannakouris N, Lopez-Miranda J, *et al.* (2008) Postprandial triacylglycerol metabolism is modified by the presence of genetic variation at the perilipin (PLIN) locus in white populations. *Am J Clin Nutr* **87**, 744–752.

34. Lee Y-M, Skurk T, Hennig M, *et al.* (2007) Effect of a milk drink supplemented with whey peptides on blood pressure in patients with mild hypertension. *Eur J Nutr* **46**, 21–27.

35. 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. *Am J Clin Nutr* **87**, 761–770.

36. 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 2010).

37. World Health Organisation, International Clinical Trials Registry Platform (2010) http://www.who.int/ictrp/en/ (accessed October 2010).

38. International Committee of Medical Journal Editors (2010) http://www.icmje.org/publishing_10register.html (accessed October 2010).

39. Krleza-Jeric K (2005) Clinical trial registration: the differing views of industry, the WHO, and the Ottawa Group. *PLoS Med* **2**, e378.

40. Bingham SA (1991) Limitations of the various methods for collecting dietary intake data. *Ann Nutr Metab* **35**, 17–27.

41. 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 2010).

42. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (1996) Guideline for Good Clinical Practice E6(R1). http://www.ich.org/LOB/media/MEDIA482.pdf (accessed October 2010).

43. US Department of Health and Human Services, Food and Drug Administration (2010) Guidance for Industry and Investigators Safety Reporting Requirements for INDs and BA/BE Studies. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM227351.pdf (Accessed 7 October 2010).

44. Hills M & Armitage P (2004) The two-period cross-over clinical trial. *Br J Clin Pharmacol* **58**, S703–S716.

45. Matthews JN, Altman DG, Campbell MJ, *et al.* (1990) Analysis of serial measurements in medical research. *BMJ* **300**, 230–235.

46. Rowe SR, Alexander N, Clydesdale F, *et al.* (2009) Funding food science and nutrition research: financial conflicts and scientific integrity. *Nutr Rev* **67**, 264–272.