REVEIW ARTICLE

Evidence-based nutrition — the impact of information and selection bias on the interpretation of individual studies

B M Margetts, H H Vorster, C S Venter

This paper in the series on evidence-based nutrition guides the nutrition scientist and practitioner regarding how to judge whether information or selection bias has occurred in reported studies. Information bias occurs when there are either random or systematic differences in the measurement of exposures, outcomes and possible confounders (effect modifiers). These variables are categorised for nutrition studies and the types of questions to ask in order to assess potential bias in obtaining this information are explained using appropriate examples. Selection bias occurs if the relationship between exposure and outcome for those who participate in the study is different to that for those from the population who are eligible to participate but do not. The paper shows how factors affecting inclusion of subjects at the beginning of the study might introduce bias and gives guidelines for appropriate sampling, indicating how knowledge of the potential sources of bias in nutrition studies should be used in the design and analysis of the study. The paper furthermore gives practical guidance on how to determine sample size, using variance, size of the hypothesised effect, level of statistical significance and power in these calculations. The paper explains the real problem of bias in nutrition studies and how to evaluate and deal with it, but also gives examples where bias may not be a problem.

In previous papers in this series we have described the general principles of evidence-based nutrition, and key concepts in the design and understanding of different types of nutritional epidemiological studies. In our first and second papers in this series we described the key questions to ask when reviewing an individual paper.[1,2] In this paper we address in more detail two important issues that often lead to bias: problems with subject recruitment (**selection bias**) and problems with the collection of information (**information bias**). If a study has a biased sample or biased data collection it will not be helpful in either guiding clinical judgment or informing a systematic review. The first steps are to decide in whom the study should be conducted and what information is required to answer the question being addressed in the study. Without a clear study question it will not be possible to judge whether either the sample selected is appropriate, or the methods used (described) are appropriate and used correctly to measure the relevant exposure, outcome and other variables. If the aim of the study is not clear and does not specifically articulate the exposure and outcome of interest, it will be virtually impossible to know whether the study has been done properly.

Public Health Nutrition, Institute of Human Nutrition, University of Southampton, UK

B M Margetts, PhD

School of Physiology, Nutrition and Consumer Science, Potchefstroom University for CHE, Potchefstroom, North West

H H Vorster, DSc

C S Venter, DSc

The aim of this paper is to give the reader guidance in how to judge whether information and/or selection bias have occurred in a paper. We provide some guidance about the key issues to consider in the way information is obtained, and in the way samples of subjects are drawn and followed up. We begin with a general discussion of bias.

Bias: General principles

Bias, defined as 'deviation of results or inferences from the truth',[3] can occur in all stages of undertaking and reporting research. Publication bias has been well reported and refers to studies that do not support the current paradigm being less likely to be published. There is even bias in the type of research questions that are asked, often driven by the source of funding.

Bias is one particular and very important type of error. There are four types of errors that can occur:

random and **systematic** errors
that can occur both
within and *between* subjects.

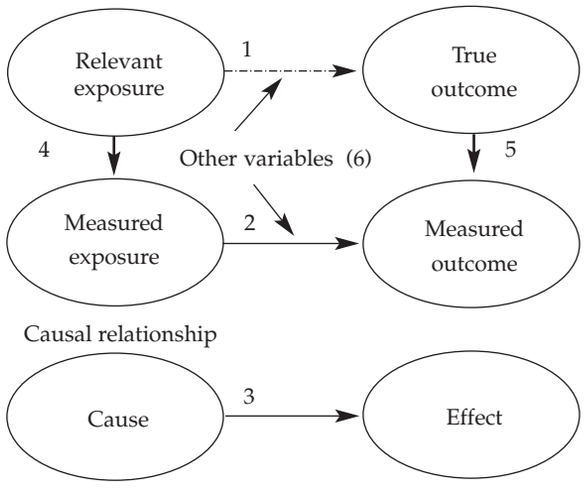
Only **systematic errors** either within or between subjects lead to bias. Random errors make it more difficult to describe the truth, and may lead to falsely rejecting a hypothesis, but do not lead to bias. Bias means you get the wrong answer. Reducing random errors does not reduce bias; bias is avoided only by reducing systematic errors. Avoiding bias must be addressed before a study starts and the impact of any potential bias on the way the study was actually conducted must be considered in the discussion.

We can define **selection bias** as bias that arises from the way subjects are included in or excluded from the study. This may include non-response bias and differential loss to follow-up in a cohort study. **Information bias** occurs when certain sorts of subjects provide different information from the rest of the sample. This may include recall or interviewer bias and will be discussed further in later sections after the basic principles of obtaining information and recruiting samples have been described.

Obtaining information: Measuring exposure, outcome and other variables

To ask any research question, information must be obtained; this information can be described as exposure, outcome or other variables. This information could be obtained by interview, by taking biological samples or physiological measurements, or by measuring health outcomes such as morbidity or mortality by specific causes. Irrespective of how the information is obtained it is important to ask whether the information obtained is appropriate to ask the question being posed by the research. By appropriate we mean *measured with the required accuracy to reflect the underlying truth that is being sought* (but which can never be known). No measure is perfectly accurate and precise, and it is therefore important to ask whether the errors that are likely to have occurred in the way the information was collected are likely to be random or systematic and therefore likely to lead to error or bias.

By convention, exposure describes those factors thought to be the causes, and outcomes are those variables that are the effect or consequence of the cause. Depending on the question being asked, a variable could in different studies be both an exposure and an outcome. For any specific question exploring a causal pathway, the exposure has to come before the outcome, and a variable cannot be both an exposure and an outcome at the same time. Fig. 1 describes the theoretical relationship between exposure and outcome and how it relates to cause and effect.



1 = the true relationship between relevant exposure and outcome.

2 = the observed relationship between measured exposure and outcome.

3 = the true causal pathway. The cause must precede the effect.

1 and 3 are the same if other variables6 (confounders) are either absent (no residual confounding) or taken account of (by stratification or mathematical adjustment).

4 = the relationship between relevant exposure and measured exposure.

5 = the relationship between true outcome and measured outcome.

6 = variables that should be measured and reported that may influence the relationship between exposure and outcome.

The extent to which the measured exposure and outcome vary from the true and relevant measures should be described in a validation study that presents the measurement errors.

Fig. 1. Relationship between exposure and outcome, cause and effect.

Nutritional exposures

These may relate to:

1. Dietary habits/food patterns and nutrient intake

- Measured
  - Consumption of food or drink (including alcoholic and non-alcoholic beverages)
  - Breast/bottle-feeding; weaning practices
  - Food and nutrition security (including access to shops, car ownership, etc.)
  - Knowledge and attitudes about food
  - Nutrient intake (derived from food composition tables or direct analysis of foods)
  - Non-nutrient intake, including biochemical modifiers of metabolism (e.g. non-nutritive antioxidants), food additives, food contaminants and toxins.
- Expression of intake
  - Average daily dose (g/day)
  - Cumulative dose (lifetime exposure, e.g. of lead).
- Derived measures or ratios
  - Nutrient density (% energy from fat)
  - Compared to a requirement
  - % above Reference Nutrient Intake or Estimated Average Requirement
  - Energy intake per kg body weight.

2. Biochemical exposures

- Levels of nutrient and non-nutrient in circulation or in tissues (concentration in blood, urinary excretion; fat biopsy or measure of flux)
- Genetic susceptibility
- Hormones
- Genetic modifiers of absorption or metabolism.

3. Anthropometry

- Measured
  - Height

- Weight
- Circumferences (e.g. waist, hip, head, chest, arm)
- Skinfold thicknesses.

- Comparison or derived measures or ratios
  - Ratios of measures
    Waist-to-hip ratio
    Weight for height (body mass index (BMI), ponderal index)
    Expressed as percentage of a standard (wasting)
    Weight for age
    % body fat from sum of various skinfolds.
  - Compared to a reference standard
    Percent above or below standard
    Z-scores
    Wasting and stunting (weight for age; weight for height).

4. Clinical measures

- Clinical signs
  - Thinning hair
  - Loss of skin pigment, flaky skin
  - Signs of specific nutrient deficiencies.

Expression of exposure measures

1. Continuous measure

- Total cumulative dose
- Average dose
- Dose at critical times/induction period
- Percent of standard.

2. Discrete measure

- Eat a particular food item (Yes/No)
- Divide continuous measure into thirds, fourths, etc. and express intakes in these groups
- Percent above/below standard.

Outcomes

Anthropometric measures are often used as both exposure and outcome (and as other variables) measures in different studies. For example, rate of growth in the first year of life could be regarded as an exposure variable in studies of risk of subsequent childhood illnesses (measles, etc.). Rate of growth in the first year of life could be an outcome variable in a study of breast-feeding practices.

Most often outcomes are disease or health status, either morbidity or mortality measures, or 'risk factors' such as hypertension, diabetes, high serum cholesterol or high homocysteine levels.

1. Disease states

- Morbidity
  - Amount or degree of illness
  - Quality of life measure
- Mortality.

2. Physiological characteristics and responses

- Risk factors for disease state
  - Blood pressure
  - Diabetes
  - Serum cholesterol.

Expression of outcome measures

It is the convention to express outcomes with reference both to the population at risk and to a defined time period, usually one year. For example, mortality is expressed as a rate of death per population at risk per year (standardised for age).

Incidence density (I) is the (Number of cases of the outcome that occur in a population during a period of time)/(Sum* for each individual in the population of the length of time at risk of getting outcome).

Prevalence (P) is the (Number of individuals having outcome at a specified time)/(Number of individuals in the population at the specified time).

Prevalence = incidence × duration.

Outcomes may also be expressed as the mean level of, for example, blood pressure or BMI or the percentage of people above an agreed level (and defined as for example hypertensive: systolic blood pressure above 150 mmHg or some other value, or obese: BMI > 30).

Other variables: Potential confounders and effect modifiers

These can be grouped as variables that have some influence on the apparent relationship between exposure and outcome variables. It is not always easy to decide what variables might be confounders or effect modifiers, but it is always necessary to consider what variables may influence the way an exposure may truly, or appear to, influence an outcome. An effect modifier modifies the effect of an exposure on an outcome, and is in the causal pathway. A confounder is associated with the outcome and independently related to the exposure, but is not part of the causal pathway. Anthropometric measures (e.g. BMI) may be a potential confounder or effect modifier (other variable) of the relationship between, for example, fruit intake and risk of stomach cancer.

Questions to ask about the description of the methods to consider potential for information bias

Before assessing whether information bias is likely, it is important to be clear whether the right measures are being used in the first place. Is it clear that the researchers have measured the relevant exposure and outcome? The relevant

* Person-years; person-time or risk time.

exposure is that which causes the outcome, measured at the time in the temporal sequence when the process leading to the outcome was begun. If the exposure is measured after this time it may be a distorted representation of the true relevant exposure. For most outcomes we do not know exactly when the process leading to the outcome began. For chronic diseases this is likely to be some time in the distant past (even before birth!). For infectious diseases the cause may be nearer the time of occurrence of the outcome. If an exposure is measured in the present, and inferences drawn backwards in time to the time when the outcome was initiated, it is important to have some sense of whether the current measure of diet (for example) reflects past diet at the time the dietary behaviour initiated the outcome.

In reading a paper, you should look for a critical discussion of whether the measure of exposure used in the study is likely to reflect the relevant exposure. Some understanding of the underlying mechanisms might help.

The 'relevant' exposure can be defined according to a number of parameters:

1. **Study type**. Ecological, cross-sectional, analytical and experimental studies require measurements made at different levels: national, community, household or individual (Table I).

2. **Time period**. Nutritional exposures can be chronic or acute in their effects. Deciding on the time at which to assess an exposure is critical to the purpose of the study. A cohort study that characterises nutritional status in terms of both dietary intakes and blood biochemistry may provide information relevant to the initiation of cancer but not necessarily to its progression.

3. **Point of measurement.** Relevant exposures can be measured in terms of food consumption, nutrient intake, blood and tissue levels of nutrient, functional consequences of nutrient action (including genetic interaction) and excretion.

4. **Type of measurement.** The list of exposures given at the start of this section provides examples of exposure measures that are direct (foods, nutrients), functional or metabolic (physiology, biochemistry), cumulative (anthropometry) or indirect (sociodemographic and cultural).

Table I. Measures of dietary exposure for different study types

Epidemiological study	Level of aggregation required; expression of information	Method
Population or household level		
Aggregate population/ecological	• Average per capita intake compared across countries or regions or households • Trends over time within a country or region or household	Food disappearance Food balance sheets Household budget surveys
Community experiment or intervention	Group level of analysis: Compare outcomes for different exposures	As above or Sentinel assessment of representative individuals
Individual level		
Cross-sectional	Absolute level Ranking	Food records 24-hour recall, FFQ Biochemical markers
Case-control	Past exposure at time of initiation or as proxy for past.	FFQ of present or past diet Diet history
Cohort	Subjects ranked and categorised Absolute accuracy not required	Food records 24-hour recall FFQ Biochemical markers
Experimental study	Individual level: usually needs to be accurate at absolute level	Food records 24 -hour recall FFQ Diet quality indices Biochemical markers

FFQ = food frequency questionnaire.

If the relevant measures have been defined, information bias can occur if there are systematic differences in the way exposure and outcome are measured in different groups within the study. There are many potential sources of information bias, but two common ones are recall and interviewer bias.

1. **Recall bias.** Any study attempting to obtain information from subjects about events in the past may be subject to recall bias. This is particularly relevant to case-control studies. Recall bias can be affected by the time interval since exposure, the degree of detail about the exposure that is required, personal characteristics of the subjects, the perceived social desirability of the exposure under investigation (for example, smokers might over-report their fruit intake), and the significance of the events under study. These biases may or may not be different for cases and controls.

2. **Interviewer bias.** Interviewer bias may occur when there is a difference in the way the information is obtained, recorded, processed, and interpreted in different groups in the study by the interviewer. If interviewers assess exposure in case-control studies and they know whether the subject is a case or a control, they may solicit the information differently. This is less likely to be a problem in a prospective cohort study where outcome is not known at the time of obtaining the exposure information.

The way subjects are followed up, or the completeness of information obtained during follow-up within a cohort or experimental study, may also introduce bias. If a distinct subset of subjects (the less healthy, the poorer, or those with high or low exposure status) are lost to follow-up, a biased result may occur.

Interviewers should be blinded as to the nature of the question under investigation and they should be carefully trained so that they collect information in a standardised way throughout.

Accuracy and precision

Two other terms that are often (wrongly) used interchangeably need to be defined to avoid confusion: accuracy and precision. Accuracy is 'the degree to which a measurement or an estimate based on measurements represents the true value of the attribute that is being measured'.[3] Precision is 'the quality of being sharply defined or stated'. For continuous variables it can be described by the standard deviation of a series of replicate measurements.[3] Precision does not imply accuracy (see boxed example alongside).

Example

A 7-day weighed record may measure an individual's fat intake more precisely than a 24-hour recall because it can characterise the within-subject random errors, which a 24-hour recall cannot. It may also be a more accurate measure of that person's true fat intake if there is no bias in the way that subject records his or her diet. But if a subject omits a range of fatty foods from the 7-day weighed record, but includes them in the 24-hour recall, the 24-hour recall may be a more accurate measure of that person's true fat intake. In most epidemiological studies the aim is to characterise the group; if all the subjects in the study leave out the high-fat foods from the 7-day weighed record, but include them in the 24-hour recall, the 24-hour recall will be a more accurate estimate of the group fat intake, even though it is considered to be a less precise measure of an individual's intake.

If it is possible to identify whether certain types of subjects will be likely to systematically alter their reported intake, it is possible to allow for this in the study design, by having enough subjects to undertake stratum-specific analyses. If this is not possible, bias may occur. For example, if overweight subjects always omit the high-fat foods from their 7-day weighed record, and the rest of the population does not, the average fat intake for the overweight compared to the rest will be an underestimate of their true intake. As long as the weight status of subjects is not then ignored in describing the fat intake of the sample, no bias will occur. If, however, the results for overweight and the rest of the sample are pooled and the overall average presented, there will be bias.

Two important points arise from this example. If you know about potential sources of bias before the study begins, this can be taken into account in the design and analysis of the study. Secondly, the impact of errors on the interpretation of a study depends on the question being asked. If an absolute measure of intake is required an absolutely accurate measure of intake is required. If subjects are to be ranked and risk of disease assessed in relation to levels of exposure (high versus low intake) absolute accuracy is not required, but the correct ranking of subjects is.

Subject selection and recruitment: Avoiding selection bias

The aims of the study dictate the best approach to drawing a sample to explore the relationship between the relevant exposure and outcome. The first question to ask is what the best way to address the hypothesis is. Most epidemiological studies are concerned with comparing and contrasting two or more groups in some measure of their exposure or outcome frequency. The objective of drawing a sample from a population is to obtain a measure of outcome such that the measure obtained in the sample is a reasonable reflection of the true outcome in the population. A bias occurs if the relationship between exposure and outcome for those who participate is different to that for those from the population

who are eligible to participate but do not. Therefore, any factors that affect the inclusion of subjects at the beginning of a study might introduce a bias. It is then a matter of judgement as to how important this bias might be in interpreting the results of the study.

Sampling strategies and potential for selection bias need to be considered for each type of study design. Study designs can be divided into observational studies (ecological, cross-sectional, case-control, cohort), where the observer does not alter exposure but simply observes it and assesses the relationship between observed exposure and outcome, and experimental studies (randomised controlled trial, intervention study), where the observer seeks to change exposure and assess what happens to outcome as a result of that change. The best way to draw a sample for each type of study design needs to be considered, mainly to ensure that selection bias is avoided and that the results are generalisble (at least to a defined population) and that the question is answered in the most efficient way possible.

A key aim of sampling is to ensure a large enough sample size in the levels of exposure across which risk of outcome is to be assessed to reduce the potential for chance findings. On the other hand a study should not be larger than is required because it wastes a lot of effort. Unless the whole population is being studied, it is usually necessary to undertake some form of sampling. The concern then is to make sure that the sampling does not lead to bias. How to judge whether the sample selected for a study is appropriate and not biased? Check to see if the following have been described:

1. Have the author/s described the source population from which the sample is drawn?

2. Have they described the process of drawing the sample in sufficient detail to assess whether any bias may have occurred?

3. Have they justified how many subjects they recruited (sample size calculation and power)?

4. Have they described response and drop-out rate?

In the design of the study some strategy should be considered as to how information on those people who refuse to participate, or who subsequently drop out, could be obtained. Even if this information is only age, gender, and occupation it will allow an estimate of whether those who participate in the study reflect the population from which they are drawn. However, a non-representative sample may not affect the internal validity of the study.

Table II presents the different questions that need to be asked for different types of study designs.

Sampling and selection bias in cross-sectional studies

Cross-sectional studies are often undertaken to describe how commonly certain characteristics occur in a population, and assess whether these characteristics are associated with outcomes (health). In other words they are looking to draw a sample in such a way that the sample reflects the relevant population. The first step in drawing such a sample is to be able to define the relevant population (devise a sampling frame) in such a way that a random sample of the population can be drawn from it. Sampling frames may be pre-existing lists of people (lists of patients in a general practice, electoral registers, school registers) or some other units (countries with food balance sheet data). Some sampling frames are geographical (e.g. list of postal codes). Alternatively, they may build up over time and only be known retrospectively (e.g. patients attending an outpatient clinic over a period of 4 months), but the rules for their construction must be made clear at the outset.

Once the sampling frame has been defined, a subset of the population (the sample) must be selected. In order for the findings from the sample to be generalisable, it is important that the sample be representative of the population from which it is drawn. The underlying principle is that each unit in the sampling frame should have the same chance or probability of being selected, i.e. the sample should be randomly drawn from the sampling frame. Selection bias is avoided by randomisation. If sub-sections of the population are deliberately excluded from the sampling frame the results cannot be generalised to that sub-section. If only sick children attend the community clinic, the estimates of illness or dietary patterns among these children will not necessarily reflect the source population. The results of the study may still be useful, but should not be over-generalised.

Different approaches to drawing a sample

A number of different approaches are used to draw a sample: simple random sampling, systematic sampling, stratified sampling, cluster sampling and multi-stage sampling. In order to draw a random sample all the units in the sampling frame need to be numbered sequentially. For simple random sampling random number tables (or a computer programme) are used to draw the required number of sampling units. The sampling fraction is the probability each unit has of being selected; for example if the required sample size is 100 and there are 1 000 units, the sampling fraction is 0.1. Unfortunately numbered sampling frames do not always exist for the target population. Systematic sampling avoids the need for a sampling frame; once the sampling fraction (say 1 in 10) is agreed then the first unit is randomly selected from the first 10 subjects; after that every 10th subject is recruited until the required sample size is achieved. Stratified sampling is used when the exposure or outcome of interest may differ between different subgroups within the sampling frame; for example weight by different age groups, or by gender. Each subgroup should be sampled separately; if the sample size is adequate this also allows for subgroup-specific estimates of exposure or

Table II. Sampling questions to be asked

Epidemiological study	Level of aggregation required; expression of information	Sampling issues: questions to ask to assess likely presence of bias
Population or household level		
Aggregate population/ecological	• Average per capita intake compared across countries or regions or households • Trends over time within a country or region or household	Consider representativeness How were countries or sampling units selected? How were years selected for inclusion?
Community experiment or intervention	Group level of analysis: Compare outcomes for different exposures	As above
Individual level		
Cross-sectional	Absolute level Ranking	What is source population? Sampling units defined? Response rate described? Differential response in sub-groups? Have key sub-groups been included? Representativeness an issue?
Case-control	Past exposure at time of initiation or as proxy for past.	Source of cases? Community or hospital controls? How controls selected? Controls from same source population?
Cohort	Subjects ranked and categorised Absolute accuracy not required	Sample selected to describe or maximise heterogeneity of exposure: has loss to follow up been described?
Experimental study	Individual level: usually needs to be accurate at absolute level	Not likely to be representative Aim to assess biological phenomena

outcome, as well as more precise population estimates. Cluster sampling samples all units from within each cluster, such as for example a family. This is a more efficient way to draw a sample, but the clustering factor needs to be taken into account in the analysis. Multi-stage sampling is a sequential series of sampling frames from which units are drawn. If the aim is to limit the geographical spread of the sample, cluster or staged sampling is appropriate. For example, there might be four stages. At each stage, a random sample is drawn of: (*i*) towns: representatives of all towns in a country or region; (*ii*) postal sectors: representatives of all postal sectors within the chosen towns; (*iii*) private addresses: representative of all private addresses within the chosen postal sectors; and (*iv*) individuals: representative of all individuals (possibly within a designated age and gender group) at those addresses.

Selection bias and response rate

If people are asked to be in a study and refuse, this needs to be documented. There is no agreed level of response rate that is acceptable; the main concern is to avoid selection bias, and even a high response rate (above 90%) may be biased if all those who refuse differ from the rest of the sample.

Sampling and selection bias in case-control studies

Case-control studies compare levels of exposure in people with (cases) and without (controls or referent group) a particular outcome. Usually cases are recruited from a clinic or centre where people with that outcome go for treatment or diagnosis. For example a case-control study may recruit stunted children attending a maternal and child health clinic, and compare their diet in the past with a group of children from the sample source population, but who at the time of the study are not stunted. Ideally the recruitment of cases should be for newly diagnosed, or incident cases, as diagnosis in the past may alter behaviour or the reporting of behaviour and therefore lead to the wrong answer. It is important to be able to define the source population from which the cases arose; in the example, this may be defined as all children within walking distance of

the clinic who would go to the clinic for routine checks. Controls are then drawn from this source population. Bias may arise if controls do no come from the source population.

A bias may also occur if the response rate is different in cases and controls. Cases, having recently been diagnosed, may be more likely to participate because they may feel that they have a vested interest in finding out more about why they got the disease. Controls, on the other hand, may not have the same concern, and those who participate may be more health-conscious or behave differently from the dynamic population they are meant to represent.

Sampling and selection bias in cohort studies

In **cohort (and experimental) studies**, where subjects are to be followed up for many years, the biggest bias that usually occurs is differential loss to follow-up. Often the sample is selected from a group of people who belong to an association or organisation that keeps track of people. For example, doctors or nurses who are required to maintain registration with their professional body are easily followed up if the researcher has access to these registration details. While doctors are certainly not typical of the entire population, if the question is to assess, for example, the effect of fruit consumption on risk of stroke, as long as there are sufficient doctors with different levels of fruit consumption (established *a priori*) and sufficient to either have or not have a stroke, it is unlikely that the relationship between fruit and stroke will be biased. The prevalence of other confounding factors may well be different, but the number and distribution of these can be taken into account in the sample size.

In a retrospective cohort study of the relationship between infant feeding practices and childhood diarrhoea, some subjects might be traced through health visitor records and some might be obtained through subjects themselves; it is likely that the latter group might be different from the group obtained from the records.

Sampling and selection bias in experimental studies

Experimental studies are often undertaken on a selected group of subjects who volunteer and are likely to comply with the experiment. In these studies the aim is usually to describe the effect of changing an exposure and assessing what effect that change has on a biological phenomenon. The experiment will be compromised if the subjects do not comply with the required change in exposure, and this should be a key criterion for selecting the subjects, as well as ensuring that the number of subjects is big enough. For these results to be generalisable it must be assumed that the biological phenomenon is universally relevant, and does not only apply to the selected group of people studied.

In experimental studies, subjects or groups in the experiment should always be randomly allocated to experimental group. The subject or the researcher should have no influence over which group the subject will be allocated to in the study. Randomisation of treatments ensures that any background or constitutional characteristics of the subjects cannot affect the effect of the treatment/intervention on the outcome. Once subjects have been randomised it is important to avoid/minimise loss to follow-up. Usually the people who are lost differ from those with whom contact is maintained, and this will then distort the measure of effect. In selecting a sample for an experiment, as in a cohort study, it may be more important to select a sample to minimise the chance of loss to follow-up. If many people are asked to be in an experiment but refuse, this will not lead to bias, as long as this refusal occurs before subjects have been randomised to treatment or control group.

Determining sample size and number of observations

How to judge whether a sample size is big enough to answer the question being addressed? The aim should be that the size of the effect *a priori* judged to be biologically important can be measured with the required accuracy.

Even if there is no information or selection bias a study may draw the wrong conclusion that there is no relationship between exposure and outcome if the sample size is too small. If a study produces statistically significant results, and is free from bias, it is likely that it is safe to conclude that the exposure is associated with the outcome. Even where statistically significant results are reported, it may not always be possible to draw causal inferences, because of the effects of unknown residual confounding, or where it is not possible to be sure about the temporal sequence, as in a cross-sectional or case-control study.

Four factors affect the required sample size: the variance (within and between subjects); the size of the hypothesised effect (difference in risk between outcomes by level of exposure); level of statistical significance (usually described as α; conventionally taken as 5% level) and power ($1-\beta$; conventionally at 80%).

Variance is important in the calculation of sample size and power. There are many factors which contribute to the variance of a quantity. Two factors in particular can be identified. The first is related to diversity *between* subjects. The second is related to diversity *within* subjects. The extent to which between- and within-subject variation affects the power of a study depends on the level of analysis and the study question being addressed. A study that requires an accurate estimate of within-subject variance (individual level of analysis) will need to take account of within-subject variation. If the aim is to describe the distribution for the population, a less precise estimate for each individual, but with a larger sample to describe between-subject variation, will be adequate.

'Statistical significance' is usually expressed in terms of *p*-values: $p < 0.05$ can be interpreted as saying that there is less than a 5% chance of observing what was observed if the null hypothesis is true. Conventionally hypotheses are written as the null hypothesis (H_o) and the alternative hypothesis (H_1); commonly the hypothesis described in a study is the alternative hypothesis. This is a bit pedantic, but strictly it is not possible to prove anything, so by default we should assume no effect (or the null), and then only reject the null hypotheses (and accept the alternative) if certain agreed *a priori* criteria apply, such as $p < 0.05$. For example, we may have a study to assess the relationship between fruit consumption and breast cancer. The null hypothesis is that there is no relationship between higher fruit consumption and lower breast cancer. We reject this hypothesis, and accept the alternative hypothesis that there is an inverse relationship between fruit intake and breast cancer when $p < 0.05$ for the test measure of effect.

When *p* is small (i.e. < 0.05), by convention we reject the null hypothesis. Remember, also, that when rejecting the null hypothesis because *p* is less than 5%, there is always the chance of making a mistake, i.e. the null hypothesis really *is* true and we have rejected it incorrectly. This is called a **type I error.** The chance of making a type I error (designated by the Greek lower case letter α) is the same as the value for *p*.

'Power' is the probability of being able to demonstrate a statistically significant finding, should one exist. Fig. 2 shows that the underlying assumption regarding "power", therefore, is that the null hypothesis is false. However, just as there was the chance of making a type I error (the incorrect rejection of a true null hypothesis), so there is the chance of making a *type II error* (the incorrect acceptance of a false null hypothesis). The greater the power in a study, the less the likelihood of making a type II error.

If a sample *is* biased, increasing the sample size will not yield results that are representative of the population, and the results may therefore be misinterpreted. A large *biased* sample size therefore does not increase the likelihood of identifying statistically significant relationships.

Concluding remarks

Bias is a very real problem in some studies such as dietary intake studies. Nutritionists need to be aware of the importance of bias and its ability to devalue otherwise well-designed studies. However, bias is not always a problem. There are situations where the bias may cancel itself out, for example, a sample surveyed on two separate occasions or a correlation study. If the bias in the method is uniform throughout the sample, and tends to reduce the estimate of intake for every occasion by some fixed amount, say 10%, then the comparison in the two examples, i.e. the change in intake between surveys or the size of the correlation, is likely to be relatively unaffected. On the other hand, bias is very serious when it affects only a subset of the sample, so that relationships between intake and other health measures become distorted. An example is when overweight subjects under-report their fat intake and underweight subjects over-report their fat intake, so disguising the association between obesity and raised energy intake.

The true situation ↓	**The decision** ↓	
	Accept Ho	**Reject Ho**
Ho is true	√ Correct decision: Probability = 1-α	× Incorrect decision: (*type I error*) Probability = α (level of significance)
Ho is false	× Incorrect decision: (*type II error*) Probability = β	√ Correct decision: Probability = 1-β POWER = (1-β) × 100

The basic formula for the number required is a function of these four elements:

$$n = \frac{2\,\sigma^2 (Z\alpha/2 + Z\beta)^2}{d^2}$$

where
σ is the variability in the observations
d is the expected outcome (e.g. the difference between two group means)
$Z_{\alpha/2}$ is the normal standard deviate relating to the chances of making a type I error (equal to 1.96, for example, when $\alpha = 5\%$)
Z_β is the normal standard deviate relating to the chances of making a type II error (equal to 0.84, for example when $\beta = 20\%$.
n is the number of observations (taken as the integer value greater than or equal to the calculated value for *n*) in *each group* being compared (*not* the total number in the study).

Fig. 2. A guideline for calculating sample size incorporating variance, size of hypothesised effect, level of statistical significance and power.

References

1. Margetts BM, Vorster HH, Venter CS. Evidence based nutrition. *South African Journal of Clinical Nutrition* 2002; **15** (2): 7-12.
2. Margetts BM, Vorster HH, Venter CS. Evidence based nutrition – review of nutritional epidemiological studies. *South African Journal of Clinical Nutrition* **15** (3): 68-73.
3. Last JM. *A Dictionary of Epidemiology*. 3rd ed. New York: Oxford University Press, 1995.

CONTINUING PROFESSIONAL DEVELOPMENT ACTIVITY FOR DIETITIANS

SAJCN CPD activity No 21 – November 2003

You can obtain 3 CPD points for reading the article: **"Evidence based nutrition: the impact of information and selection bias on the interpretation of individual studies"** and answering the accompanying questions.
This article has been accredited for CPD points for dietitians. (Ref number: 03/3/104/12)

HOW TO EARN YOUR CPD POINTS

1. Check your name and HPCSA number.
2. Read the article and answer all the questions.
3. Indicate your answers to the questions by coloring the appropriate block(s) in the cut-out section at the end of this questionnaire.
4. You will earn 3 CPD points if you answer more than 75% of the questions correctly. If you score between 60-75% 2 points will be allocated. A score of less than 60% will not earn you any CPD points.
5. Make a photocopy for your own records in case your form is lost in the mail.
6. Send the cut-out answer form **by mail,** NOT BY FAX to: SASPEN Secretariat, SAJCN CPD activity No 21, c/o Department of Human Nutrition, PO Box 19063, Tygerberg, 7505 to ***reach the office not later than 18 February 2004.*** Answer sheets received after this date will not be processed.

PLEASE ANSWER ALL THE QUESTIONS

1. Systematic errors within or between subjects lead to bias in a study.
   [a] True
   [b] False

2. Bias occurs if overweight subjects under-report their fat intake and underweight subjects over-report their intake.
   [a] True
   [b] False

3. Selection bias occurs when there are differences in the way exposure and outcome are measured in the study group.
   [a] True
   [b] False

4. Bias occurs if the relationship between exposure and outcome for those subjects who participate in a study is different to that of the rest of the population.
   [a] True
   [b] False

5. A variable can be both an exposure and an outcome at the same time.
   [a] True
   [b] False

6. The relevant exposure is that which causes the outcome, measured at the time when the process leading to the outcome was initiated.
   [a] True
   [b] False

7. Outcomes are usually expressed with reference to the population at risk as well as a defined time period.
   [a] True
   [b] False

8. Confounders have some influence on the apparent relationship between exposure and outcome.
   [a] True
   [b] False

9. Confounders are associated with the outcome and independently related to the exposure, but are not part of the causal pathway.
   [a] True
   [b] False

10. Bias may occur if the response rate is different in cases and controls.
    [a] True
    [b] False

11. The biggest bias in cohort studies where subjects are followed up for many years is
    [a] Differential loss to follow-up
    [b] Prevalence of other confounding factors

12. If a sample is biased, increasing the sample size will yield results that are representative of the population.
    [a] True
    [b] False

✂ Cut along the dotted lines and send to: SASPEN Secretariat, SAJCN CPD activity **No 21,** c/o Department of Human Nutrition, PO Box 19063, Tygerberg, 7505 to ***reach the office not later than 18 February 2004***

HPCSA number: DT |__|__|__|__|__|__|__|

Surname as registered with HPCSA: _____ Initials: _____

Postal address: _____

_____ Code: _____

Full member of ADSA: |__| yes |__| no If yes, which branch do you belong to? _____

Full member of SASPEN: |__| yes |__| no Full member of NSSA: |__| yes |__| no

"Evidence based nutrition: the impact of information and selection bias on the interpretation of individual studies"
BM Margetts, HH Vorster, CS Venter

Please color the appropriate block for each question
(e.g. if the answer to question 1 is a: 1) a ■ b)

1) [a] [b]	2) [a] [b]	3) [a] [b]	4) [a] [b]
5) [a] [b]	6) [a] [b]	7) [a] [b]	8) [a] [b]
9) [a] [b]	10) [a] [b]	11) [a] [b]	12) [a] [b]