Ministry for Primary Industries
Manatū Ahu Matua

Systematic review of a food-health relationship

Guidance document on how to self-substantiate a food-health relationship in order to make a new general level health claim

Prepared by:
Biosecurity Science, Food Science & Risk Assessment Directorate,
Regulation and Assurance Branch

February 2016

About this document

This document is intended to provide industry with guidance on how to self-substantiate a food-health relationship in order to make a new general level health claim.

The guide includes detailed information and practical examples for establishing a causal relationship between a food, or a property of a food, and a health effect.

The guide is based on the legal requirements that relate specifically to self-substantiation of health claims which are set out in Standard 1.2.7 Nutrition, Health and Related Claims. While guidance material is not legally enforceable, stakeholders are encouraged to discuss significant departures from this best practice approach with MPI to avoid expending resources on the development of alternative approaches.
Please note that the examples described in this document are used purely for illustrative reasons and are not exhaustive.

This publication is available on the Ministry for Primary Industries website at
http://www.mpi.govt.nz/news-resources/publications.aspx

Contact details

Ministry for Primary Industries
Regulation& Assurance Branch
PO Box 2526
Wellington 6140
Email: healthclaims@mpi.govt.nz

Disclaimer

Table of Contents

1 Executive summary

Standard 1.2.7 Nutrition, Health and Related Claims provides a route for the food industry to self-substantiate a food-health relationship in order to make a general level health claim. Self-substantiation requires a systematic review supporting a causal association between the food and the health effect. The purpose of this guidance document is to provide industry with more detailed information and examples that will assist in establishing a causal relationship between a food, or a property of a food, and a health effect by the process of a systematic review according to the requirements outlined in Standard 1.2.7.

A systematic review is an overview of a specific research question that systematically identifies, appraises, selects and synthesises all high quality primary research evidence relevant to that question. An essential component of the systematic review is to assess the totality of the relevant scientific evidence regardless of the study results. In conducting a systematic review for the purpose of substantiating a food-health relationship, the food or the property of the food and the health effect that are the subject of the systematic review must be defined. From this, the review question should be developed so that the search strategy and the inclusion/exclusion criteria for the systematic review can be formulated. When executing the search strategy and filtering through the references, it may be useful to use bibliographic or systematic review software to help manage the list of references and complete certain parts of the systematic review. Once the list of all relevant studies identified through multiple sources has been finalised, the evidence must be evaluated. Constructing tables by extracting data and information on key components of each of the studies included in the systematic review is an essential part of the requirements and it will assist in assessing the quality of the studies, evaluating the consistency of results, and for considering the generalisability of the findings. It is important that the systematic review includes evidence from studies that are high quality (preferably double blind placebo-controlled randomised trials) because this will help to improve the confidence in the conclusion reached by the systematic review. Certain aspects of study quality must also be assessed and completing a study quality appraisal tool will help to evaluate specific components of study quality such as bias. Deciding whether there is a causal relation between the food, or the property of the food, will include an assessment of the consistency, strength, dose-response and temporality of the results from high quality studies. It is also important to consider whether there is a biologically plausible mechanism for the effect of the food on health and the amount of the food or the property of the food that will achieve the proposed health effect in relation to what is achievable in an Australian and New Zealand diet.

Deciding whether there is a causal relationship between a food or a property of a food involves a certain degree of judgement and this guidance document is meant to provide industry with some insight to the processes used by the Ministry for Primary Industries (MPI) to judge this.

2 Purpose of this guidance document

This document is meant to act as a supplement to the guidance document written by Food Standards Australia New Zealand (FSANZ) 'Guidance on establishing food-health relationships for general level health claims' which provides important information to guide food businesses in establishing a causal relationship between a food or property of food and a health effect (food-health relationship) by a process of systematic review for the purpose of making a general level health claim. MPI's document contains more detail and in particular provides more examples of each of the requirements of the systematic review outlined in Schedule 6 of Standard 1.2.7. This (MPI's) guidance document should also be used in conjunction with the relevant sections of the document from the Implementation Subcommittee for Food Regulation (ISFR) health claims working group 'Getting Your Claims Right - A guide to complying with the Nutrition, Health and Related Claims Standard of the Australia New Zealand Food Standards Code'.

This document goes through in detail each of the required elements of a systematic review which are outlined in Schedule 6 of Standard 1.2.7 Nutrition, Health and Related Claims. In doing so it will mean that all components of a systematic review as outlined in Figure 1 below are covered.

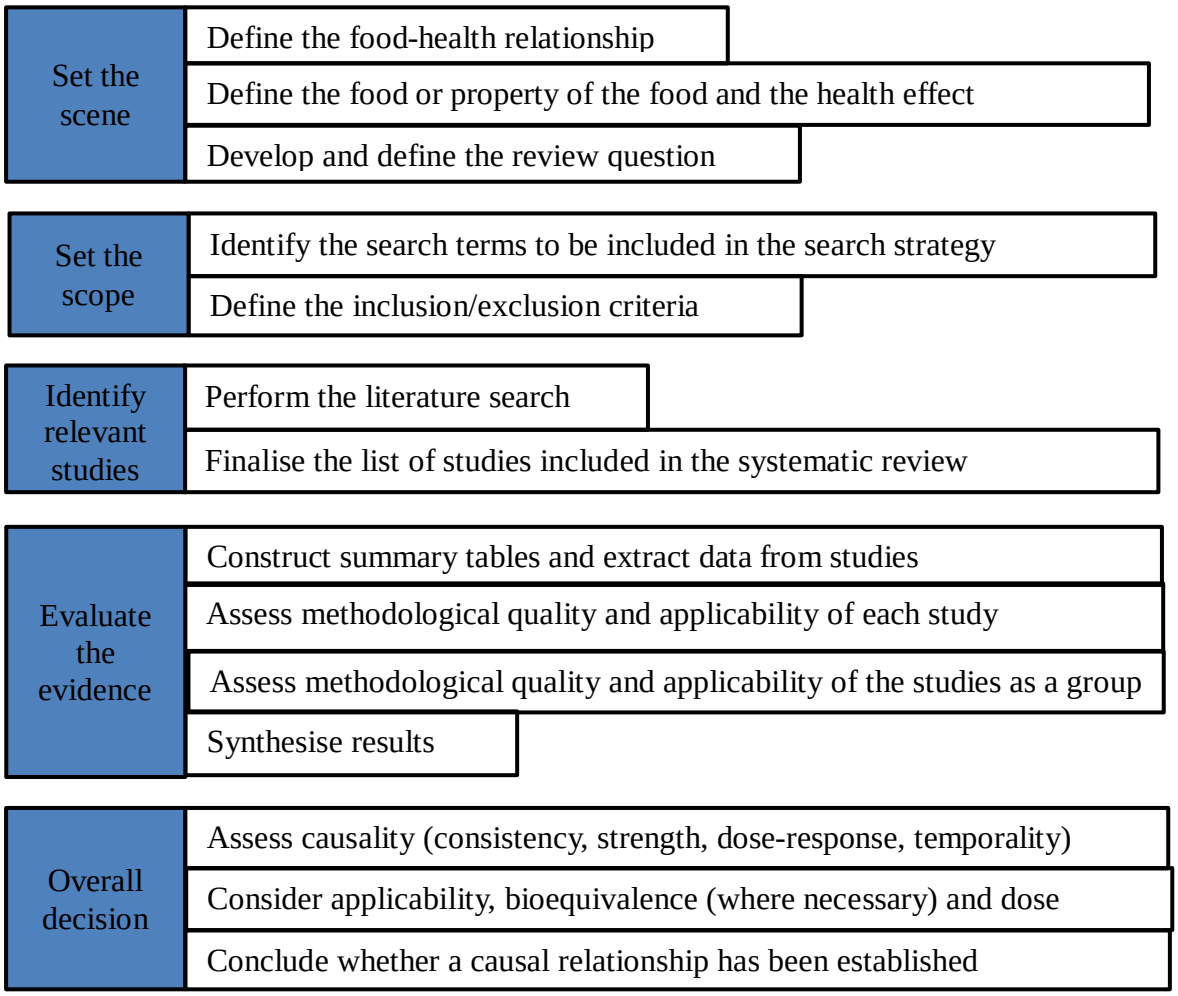

Figure 1. Overview of the process for conducting a systematic review to self-substantiate a food-health relationship

If a NZ food business is considering completing a systematic review to self-substantiate a food-health relationship, MPI would encourage the food business to make contact with MPI who will be able to provide you with regulatory and technical advice well in advance of notification stage. We will treat confidential information you provide in relation to any guidance we give with complete confidence. Please email MPI on health.claims@mpi.govt.nz.

3 Background information on a systematic review

A systematic review of a food-health relationship forms the basis of what is required to make a new self-substantiated general level health claim. The purpose of a systematic review is that it allows for individual studies to be interpreted in the context of other similar studies rather than considering the results of studies in isolation. In order to produce reliable results, the process of conducting a systematic review includes methodically locating, critically appraising and synthesising the scientific evidence.

In adequately nourished populations, the health effects of certain foods or the properties of foods such as nutrients are likely to be moderate. Therefore, in order to detect these moderate effects care must be taken to ensure that biased comparisons do not lead to the conclusion that there is a causal effect of a food on health when one does not exist. An example of a biased comparison is if participants who are judged as non-compliant with the intervention are excluded from the analysis, this will create an imbalance between the groups in the study. This imbalance will be problematic where the reasons for not complying are related to a lack of effect or adverse effects of the intervention.

The conclusions reached from a systematic review of the effect of a food or property of a food on an aspect of health will depend upon the processes used by those conducting the systematic review; for example, the relationship that the systematic review addressed, search strategy used, inclusion/exclusion criteria, method used to categorise high quality studies, and interpretation of the evidence from the studies included in the systematic review. An important factor to judge in this process is the extent to which the systematic review has included all relevant studies and successfully reduced the amount of bias (Chalmers 2003). Even so, there may be a bias in which studies get published and so it is important to assess whether even a systematic compilation of literature yields the correct overall view of the relationship.

3.1 Guidance on the Ministry for Primary Industries' evaluation of the systematic review

The requirements for the systematic review to support a food-health relationship for a general level health claim are laid out in Schedule 6 of Standard 1.2.7 – Nutrition, Health and Related Claims.

Upon receiving information that a self-substantiated food-health relationship has been notified to FSANZ by a New Zealand food business, MPI will assess the notified relationship to see whether it needs to be investigated further.

Clause 19(1) (d) of Standard 1.2.7 states that a person who gives the notice is required to:

> (d) if requested by a relevant authority, provide records to the relevant authority that demonstrate that –
>
> (i) the systematic review was conducted in accordance with the process of systematic review described in Schedule 6; and
> (ii) the notified relationship is a reasonable conclusion of the systematic review.

MPI is the relevant authority in New Zealand and may ask to see the systematic review to evaluate whether the systematic review meets the requirements outlined in Schedule 6 and that food-health relationship is a reasonable conclusion of the systematic review. If MPI does ask to see the systematic review, we will go through and evaluate specific components including the methods used to conduct the systematic review and the critical appraisal of studies included in the systematic review to ensure that the scientific evidence supports a causal relationship between the food or the property of the food and the health effect.

There are elements in Schedule 6 which rely heavily on the interpretation of methods and results of studies included in the systematic review in order to establish whether there is a causal relationship between a food or property of the food and a health effect. There are a number of different factors that underpin a causal relationship and to a certain extent, the process of evaluating whether the food or the property of the food causes the health effect is a matter of judgement. There may be cases where the judgements made about the evidence by those who conducted the systematic review and those made by MPI when evaluating the systematic review will differ. In order to make the process of evaluation as transparent as possible, the additional guidance provided in this report will help ensure that the judgements made about the evidence by MPI are consistent, reasoned and clear.

4 Set the scene for the systematic review

4.1 Develop the food-health relationship

In order to get started on a systematic review, the particular question needs to be developed, as this will help guide the search strategy to find relevant studies to include in the systematic review. The review question relates to the food-health relationship which describes the food or property of the food and the health effect. The food-health relationship underpins the general level health claim and the list of pre-approved food-health relationships for making general level health claims are listed in Schedule 4—5 (in the revised Food Standards Code effective 1 March 2016). It is important that the proposed direction of the relationship between the food or the property of the food is mentioned e.g. Phytosterols, phytostanols and their esters reduces dietary and biliary cholesterol absorption.

4.2 Define the food or the property of the food and the effect on health

Required elements of a systematic review - Schedule 6—2 (a):

> A description of the food or property of food, the health effect and the proposed relationship between the food or property of food and the health effect.

The exact review question will depend upon a number of different factors such as whether the health claim is about a specific food product, a nutrient in that food, or another property of the food. It is possible that a food business has to complete several research scoping exercises to assess the scientific evidence for the food or the property of the food before defining the exact nature of the food that is the subject of the systematic review. For example, it could be prunes, dietary fibre from prunes, or some other property of prunes such as the sugar alcohol sorbitol that is responsible for a proposed health effect. This will also determine which relationship needs to be assessed and what the comparator food/property will be. It should be noted that in some cases, studies will not be able to provide evidence for the exact property that causes the health effect.

It is important that the food or the property of the food is defined and characterised well. For instance, for a health claim on prunes, then it would be useful to define the prunes as dried plums of “prune” cultivars (*Prunus domestica* L.). The methods for measuring most nutrients in foods are well established and for total dietary fibre the prescribed methods of analysis are outlined in Standard 1.2.8 ‘Nutrient Information Requirements’. It would also be useful to state whether the food is consumed raw or processed further (and the method of processing). Also, in the case where more than one part of a food could be eaten, then the part that is the subject of the food-health relationship should be specific (e.g. root, leaves, flesh, skin).

An example for a property of a food would be the food-health relationship “whey protein has an effect on growth or maintenance of muscle mass” then intervention studies demonstrating an effect of whey protein compared to some other source(s) of protein that are similar in terms of the nutritional composition, especially protein and energy (e.g. casein or soy), on measures of muscle mass could be used as evidence to support this food-health relationship. Studies that provide evidence for this health effect may include those that have used whey protein in the final food product e.g. in a protein drink and studies that have used whey protein in isolation e.g. as a supplement (both types using the appropriate comparator). As with the use

of any studies that administer a supplement, there will have to be additional evidence to show that a similar effect on health is demonstrated for the whey protein when present in food. That is, demonstrating the relative bioequivalence of the property of food when consumed in the food matrix (see the FSANZ guidance document for more information). In this example, the whey protein fractionate may also need to be characterised further in terms of the methods used to separate it.

Certain properties of foods such as an "antioxidant" will need to be characterised further before forming part of a food-health relationship, defining the measurable food component(s) that have antioxidant properties, for example vitamin C.

In defining the effect on health, there are a series of documents that have been developed by the European Food Safety Authority (EFSA) Panel on Dietetic Products, Nutrition and Allergies (NDA) that detail useful information on the scientific requirements related to specific health effects that the NDA panel uses. It might also be helpful to consult someone with an area of expertise in the field of the health claim so they can advise on a suitable health effect, appropriate validated methods for measuring the most appropriate health outcomes for that effect, and a suitable study duration for being able to demonstrate the effect of the food or the property of the food on health.

Under the Food Standards Code, a health effect is defined as the following:

> **health effect** means an effect on the human body, including an effect on one or more of the following –
>
> (a) a biochemical process or outcome;
> (b) a physiological process or outcome;
> (c) a functional process or outcome;
> (d) growth and development;
> (e) physical performance;
> (f) mental performance;
> (g) a disease, disorder or condition.

Throughout this guidance document, an effect on the human body will be referred to as a health effect, an effect on health or a health outcome. Note that in the Food Standards Code, general level health claims are defined as the following:

> **general level health claim** means a health claim that is not a high level health claim.

It is what is left over after the following definition for a high level health claim:

> **high level health claim** means a health claim that refers to a serious disease or a biomarker of a serious disease.

Where a biomarker and a serious disease have the following definitions:

> **biomarker** means a measurable biological parameter that is predictive of the risk of a serious disease when present at an abnormal level in the human body.

> **serious disease** means a disease, disorder or condition which is generally diagnosed, treated or managed in consultation with or with supervision by a health care professional.

Therefore, for the purposes of making a general level health claim the following examples of health effects are given. Note, that some of these health effects may fall under more than one category.

A **biochemical process or outcome** could refer to an effect on homocysteine metabolism that is, the transfer of a methyl group to homocysteine by methionine synthase to form methionine. This could be assessed by measuring plasma concentrations of homocysteine.

A **physiological process or outcome** could refer to an effect on vision. This could be assessed by using validated tests of visual acuity (e.g. Glasgow acuity card method) and contrast sensitivity (e.g. Pelli-Robson contrast sensitivity chart).

A **functional process or outcome** could refer to an effect on skin barrier function (i.e. reducing the risk of skin dehydration). This could be assessed by measuring trans epidermal water loss using validated methods.

Growth and development could refer to an effect on children's height using the appropriate growth curves/charts.

Physical performance could refer to an effect on endurance capacity (exercising to fatigue). This could be assessed by using a validated endurance performance test such as a maximal incremental exercise test.

Mental performance could refer to an effect on attention or concentration. This could be measured by using the 'Test for Attentional Performance' which is a standardised, validated test battery (set or series of related tests) that assesses a range of specific attentional performances.

A **disease, disorder or condition** could refer to an effect on the risk of a common cold (a non-serious disease). This could be assessed by measuring the occurrence of the common cold.

Under Standard 1.2.7—8 Claims are not to be therapeutic in nature:

> A claim must not –
>
> (a) refer to the prevention, diagnosis, cure or alleviation of a disease, disorder or condition; or
> (b) compare a food with a good that is –
>
> (i) represented in any way to be for therapeutic use; or
> (ii) likely to be taken to be for therapeutic use, whether because of the way in which the good is presented or for any other reason.

Therefore, for a general level health claim related to the common cold, it should be clear that the effect on health is a reduction in the risk of the common cold, which is phrased in a similar way to the food-health relationship for high level health claims that refer to a reduction in the risk of a serious disease (e.g. calcium and vitamin D reduces the risk of osteoporosis). The claim cannot refer to the prevention of the common cold which, under the conditions outlined above, would be deemed a prohibited therapeutic claim. In deciding whether a health claim could potentially be therapeutic, it might be useful to identify whether

the claimed effect on health comes before or after the onset of the condition. Health claims that refer to the easing of symptoms of a condition such as a cold (i.e. health effects that occur after the onset) would be considered therapeutic.

4.3 Develop the review question

Once the food or the property of food and the health effect has been established, the next step is to formulate the review question so that the search strategy for the systematic review can be developed. The guidance document on conducting a systematic review by FSANZ mentions the use of Participants, Intervention, Control, Outcome (Time, Study Design) PICO(TS) to develop a review question and provides guidance on formulating the inclusion/exclusion criteria (which can be based on several aspects of PICOTSs). The following is an example review question developed for a systematic review to examine the food-health relationship of the consumption of "phytosterols, phytostanols and their esters reducing dietary and biliary cholesterol absorption":

Population/participants – generally healthy adults (with or without high blood pressure/cholesterol levels)

Intervention/**E**xposure – phytosterols, phytostanols and their esters (plant sterols, stanols and their esters) e.g. sitosterol and campesterol (the most common form found in foods).

Comparison/control – placebo control, e.g. the food without the plant sterols/stanols.

Outcome – dietary and biliary cholesterol absorption, calculated by measuring faecal cholesterol excretion

Time – 4 weeks

Study – randomised controlled trials.

It is worth noting that for a systematic review of a food-health relationship for a general level health claim, the most reliable source of evidence will come from randomised controlled trials. This is the best way to ensure that bias in the association of the food or property of the food with the health effect is kept to a minimum (Chalmers *et al* 2014). Evidence from other types of studies – non-randomised controlled trials, cohort studies and case-control studies – can also be considered when judging whether a causal relationship exists but a greater weighting will be given to randomised controlled trials, especially those that are high quality and have a large sample size. In the field of nutrition, the majority of cohort studies are used to assess the effects of diet on the risk of developing certain types of non-communicable (and often serious) diseases. Many health effects that are appropriate for making general level health claims are often measured at the same time as dietary intake is assessed in these studies. Such cross-sectional analyses does not provide reliable evidence for supporting a causal food-health relationship. The majority of case-control studies have enrolled cases with a serious disease (e.g. breast cancer) and controls without that specific serious disease. Therefore, these studies might not have assessed a health effect that that is appropriate to form the basis of a food-health relationship for making a general level health claim applicable to the healthy population. Dietary intake in large cohort studies is often collected using a Food Frequency Questionnaire (FFQ) which assesses the intake of certain foods, often collects data at a grouped level (margarine) as opposed to individual foods (margarines containing

phytosterols), and even if detailed recalls have been used, these may have been collected at a time before the new property being assessed was developed.

5 Set the scope of the systematic review

5.1 Define the inclusion/exclusion criteria (PICOTS)

Required elements of a systematic review - Schedule 6—2 (b):

> A description of the search strategy used to capture the scientific evidence relevant to the proposed relationship between the food or property of food and the health effect, including the inclusion and exclusion criteria.

Before beginning to search for relevant studies, the inclusion/exclusion criteria should be developed so that reasons for excluding certain journal articles or studies are made before viewing the articles that are retrieved from the search strategy. It is important to note that any restrictions to this criteria are justifiably sound. There must be a biologically plausible, or quality driven reason for limiting studies to a particular age group (within the adult group but not for the division between children and adults), ethnicity or sex (Higgins and Green 2009). An essential component of the systematic review is to assess the totality of the relevant scientific evidence regardless of the study results and therefore, it is recommended that studies are not excluded based on the results not being statistically significant, as this may bias the results of the systematic review. If essential information such as data on the main outcomes or the standard deviation/error is not reported in a relevant study, then it is a good idea to contact the corresponding author of the article to see whether they can provide this information.

Here is an example of exclusion criteria that were used for a systematic review of phytosterols and cholesterol absorption. Note that most of the exclusion criteria relate to aspects of the study question that were outlined using PICOTS.

1. Plant sterols are not one of the intervention or exposures
2. No placebo control group
3. Did not measure cholesterol absorption
4. Study duration is not long enough to see a sustained effect (adequate duration specified)
5. Not original research (that is, reviews, editorials, letters to the editor).
6. Not in healthy human adults

For most health effects related to general level health claims, there may be several different methods used to measure the specific health outcome. It is important to consider the extent to which the methods capture the intended effect on health, and whether they are valid and reliable measures of the health effect. Emphasis is best placed on the most valid and reliable measures in the subsequent discussion on consistency of effect, but these should be documented prior to the search undertaken.

5.2 Identify the search terms to be included in the search strategy

In generating the search strategy for a systematic review, it is important to get a good balance between including search terms that are relevant and comprehensive but are not overly exhaustive so as to generate search results containing tens of thousands of irrelevant titles and abstracts (Higgins and Green 2009). Librarians at Medical School/Health Science libraries have expertise in helping to come up with the relevant search terms and it is recommended that you seek their help in developing your search strategy once the review question has been finalised.

There will be two main components included in the search strategy: terms to search for the food or property of the food and terms to search for the health effect of interest (Higgins and Green 2009). The following is some guidance for conducting a search using PubMed, which searches MEDLINE – a bibliographic database of life sciences and biomedical information. In developing the search terms to use for the search strategy, it would be useful to search the Medical Subject Headings (MeSH) terms in the MeSH library to find the relevant subject content in journal articles to be used within the search strategy. MeSH terms are the National Library of Medicine's controlled vocabulary thesaurus used for indexing articles. The MeSH terms have a hierarchical structure so that they can be searched at various levels of specificity. When doing a search using the MeSH term "phytosterols" there are a number of subheadings that relate to a range of particular aspects of that subject (e.g. "chemistry" and "metabolism") and it is possible to select only the relevant subheadings to be included in the search. There may also be a number of other subject headings listed below the search term in the hierarchical structure. If all of the subject headings below the MeSH term are included in the search, then this term has been 'exploded'. If you do not want to include the more specific subject headings below the MeSH term then you can check the option "Do Not Explode this term".

As MeSH terms are attributed manually to specific articles, certain research articles may be missed in the attribution of the term. Therefore, it is also recommended that a search using MesH terms is combined with a free text search. Use of the search builder in PubMed means that search terms can be combined using the connectors "AND", "OR" and "NOT" so a search for "phytosterol*" under "Title/abstract" or "Text" can be combined with a search for the MeSH term "phytosterols" using "OR" for maximum coverage of the appropriate articles. It is also possible to restrict the search strategy to human studies, which will limit the number of irrelevant studies and this can be done by using a filter for humans. For more information on how to conduct a search in PubMed, please refer to the help section on the PubMed website.

Please note that the development of a search strategy is an iterative process, and search terms are usually modified based on the relevance of the articles that are retrieved, so it may take some time before the final search strategy is developed. Furthermore, this is an example of the way a search strategy would be conducted in the database MEDLINE using the portal PubMed. The way that the search strategies are executed in other databases will differ to that of PubMed and some, such as Embase, use a different system from MeSH terms for indexing articles.

Table 1. Example of search terms for developing a search strategy for the food-health relationship "Phytosterols, phytostanols and their esters reduce dietary and biliary cholesterol absorption"

PICOTS	MeSH/keywords		Text words
Participants	"humans"		
AND			
Intervention	"phytosterols"		"phytosterol*" OR "plant sterol*" OR "plant stanol*" OR "phytostanol*" OR "sitosterol*" OR "sitostanol*" OR "campesterol*" OR "campestanol*" OR "stigmasterol*" OR "brassicasterol*"
Control (if required)	Not included in this search strategy		
AND			
Outcome	Faecal cholesterol excretion "cholesterol"	OR	"cholesterol"
Timeframe	Interventions of > 3 weeks	OR	"3 weeks" OR"21 days"
Study	"randomized controlled trial" (publication type) OR "controlled clinical trial" (publication type)	OR	"randomized" OR "placebo" OR "randomly" OR "trial" OR "groups"

Allows for various forms of the word to be identified in the search e.g. phytosterol will pick up phytosterol and phytosterols.

6 Undertaking the literature search

6.1 Databases

The FSANZ guidance document lists some of the relevant databases to search and suggests that it is good scientific practice to search at least two of these databases. Some of the databases (e.g., Scopus and Embase) also search a number of conference proceedings. The way that the search strategies are executed in the various databases will differ. Another example given in the FSANZ guidance document – the Cochrane Central Register of Controlled Trials (CENTRAL) – is a bibliographic database of randomised controlled trials. It is recommended that clinical trial registry databases are also searched in an attempt to ascertain whether there are any relevant trials that have not been published. This is because a number of completed trials are not eventually published or take a long time to be published. Moreover, there is empirical evidence which indicates that studies which have results that are not statistically significant or 'negative' are less likely to be published or disseminated (Song *et al* 2000), which could bias the results of a systematic review. This method also helps to identify key studies in progress, to indicate when an appropriate time to update the systematic review might be, in order to incorporate the results of these trials.

The strategy used for searching clinical trial databases for relevant studies is quite different from that when searching other databases and so the search strategies will need to be modified accordingly. The process of searching for relevant studies in the clinical trials database should not be an onerous task given that there are far fewer studies included here than in electronic databases. If a registered trial has already been published, then there should be a reference to the paper in the clinical trial database. It is also recommended that the reference lists from included and relevant articles are searched through to find other articles that may have been missed in the database searches.

6.2 Bibliography

All the search results from the electronic databases can be exported to a bibliography and database manager such as (but not limited to) RefWorks, Endnote, Mendeley, ProCite or EPPI-Reviewer. This is an easy way of keeping track of the number of journal articles that are identified in the full literature search. These programs can usually automatically identify duplicates and it is also possible to keep copies of the abstracts (and sometimes the full text article) in these database managers. Some of these bibliography and database managers have been developed for the purpose of conducting a systematic review and have additional features such as being able to assign which criteria applies to the studies that are excluded, and there is also the possibility to conduct a meta-analysis using the software within some of the databases (if it is appropriate to do so).

6.3 Duplicate search recommended

It is recommended that at least two people complete the search strategy and the process of filtering the articles to come up with a final list of studies to be included in the systematic review. Any disagreements between reviewers can usually be resolved by discussion or the input of a third person.

6.4 Use of bibliographic and systematic review software

All the search results from the electronic databases can be exported to a bibliography software such as (but not limited to) RefWorks, Endnote, Mendeley and ProCite or systematic review software such as EPPI-Reviewer, RevMan and DistillerSR®. This is an easy way of keeping track of the number of journal articles that are identified in the full literature search. These programs can usually automatically identify duplicates and it may be possible to store copies of the abstracts and the full text articles in them.

The various systematic review software have been developed for the purpose of conducting a systematic review and have additional features such as being able to assign which criteria applies to the excluded studies and to assess the risk of bias of the studies included in the systematic review. Some of the systematic review softwares can be adapted to suit the protocol of the systematic review but some, like RevMan cannot be modified as this is the software used for preparing a Cochrane Review. Most of the systematic review softwares include the tools necessary to conduct a meta-analysis of the overall effect (if it is appropriate to do so).

6.5 Identify relevant studies

Sometimes it is difficult to know whether to include a study but it is important that all possible information on the causal relationship is included in the systematic review. It is possible to consider the association of the food or property of the food with the effect on health according to different lifestyle factors of the participants but these should be set out in the protocol for the systematic review rather than being *post hoc* analyses.

In the context of randomised controlled intervention studies, a control group refers to a group of participants who are alike in as many ways as possible to the intervention group. Both groups of participants come from the same source of participants who have been assigned to each group by random, preferably masked, allocation. One of the main points of having a control group within an intervention trial is to test what would happen over the duration of the study if they did not receive the intervention. The main outcome of this study would be a comparison of the health effect between the groups, that is; the intervention versus control. Many methods of allocating subjects to groups which are thought to be random (e.g. by alphabetical order, sealed envelopes) are not, in fact, random or are easily subverted.

Given that very few food composition tables will contain accurate information on the quantity of phytosterols, it is unlikely that there will be observational studies that have examined the association between phytosterols and cholesterol absorption, let alone with adequate control for confounding variables. Therefore, the most relevant information to address this review question will come from intervention trials.

It is a good idea that when going through the list of potential references for inclusion in the systematic review that there is a record kept of the reasons why the article was not included in the final list of studies. Often an article will fulfil more than one of the exclusion criteria so it is suggested that a hierarchy is applied when assigning an exclusion criterion to an article. For example, if a report is excluded on the basis of meeting exclusion criteria 1, 3 and 5, then exclusion criteria (1) is assigned to that citation and if exclusion criteria 3 and 4 apply, then exclusion criteria (3) is assigned to that citation. It is up to the reviewers to decide what hierarchy to use if they decide to use one.

6.6 Finalise the list of studies included in the systematic review

Required elements of a systematic review - Schedule 6—2 (c):

> A final list of studies based on the inclusion and exclusion criteria. Studies in humans are essential. A relationship between a food or property of food and the health effect cannot be established from animal and *in vitro* studies alone.

It is generally not necessary to document the reasons why reports are excluded when initially reviewing the titles or abstracts; however, this might be useful for ensuring consistency and transparency in the process, when deciding which studies to include in the systematic review. This will also depend on the number of reports screened. However, in order to meet the requirements of Schedule 6 it is *essential* that after reviewing the full-text articles the reasons for excluding reports from the final list of studies are documented. This provides justification in a transparent manner, as to why individual studies that have reached this phase of the process are still deemed inappropriate for analysis (Altman and Bland 1998).

Part of establishing a causal relationship between a food or property of a food and a health effect as outlined by Schedule 6 is also considering whether the proposed relationship is biologically plausible. Studies that might contribute to providing evidence for a biologically plausible relationship could be amongst those excluded from the final list of studies. For example, studies that were excluded because the duration was too short (e.g. the studies that examined only the post-prandial effects) might provide some useful information on the biological mechanism(s) that underlie the association being reviewed. If a hierarchy of exclusion criteria is applied, then it will be easier to locate these studies when considering the biological plausibility of the food-health relationship, later in the review process.

To summarise the processes described above, a flow diagram of studies included in the evaluation of the evidence is a useful addition to the report (Figure 2). Information specific to the health outcomes and food property being tested would need to be included.

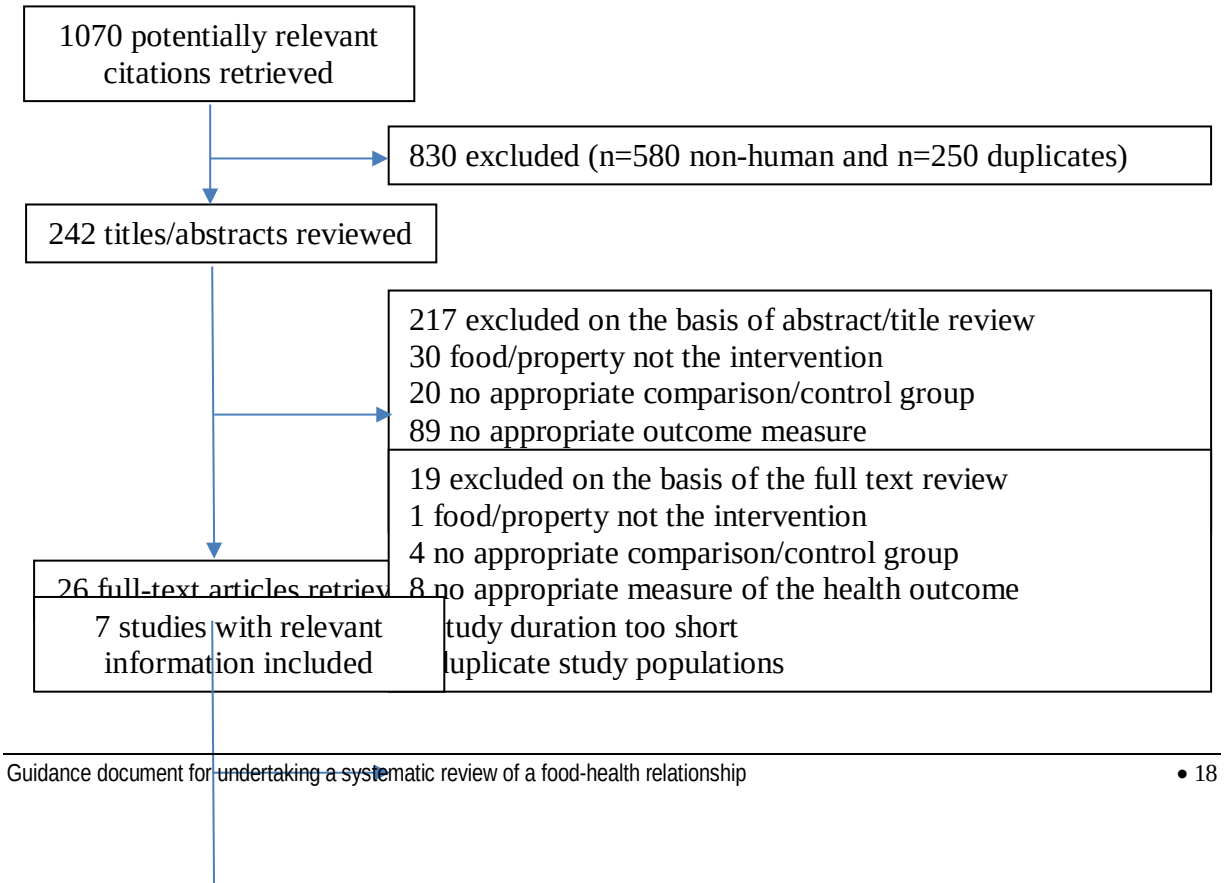

Figure 2. An example flow diagram of studies included and excluded in a systematic review process.

7 Evaluate the evidence

7.1 Construct summary tables and extract data from the studies

Once the list of studies to be included in the systematic review has been finalised, the key information from each of the studies must be tabulated.

Required elements of a systematic review - Schedule 6—2 (d):

> A table with key information from each included study. This must include information on:
>
> (a) the study reference
> (b) the study design
> (c) the objectives
> (d) the sample size in the study groups and loss to follow-up or non-response
> (e) the participant characteristics
> (f) the method used to measure the food or property of food including amount consumed
> (g) confounders measured
> (h) the method used to measure the health effect
> (i) the study results, including effect size and statistical significance
> (j) any adverse effects.

Note that this is the essential information, and it is suggested that reviewers add other items to the table that they think may be relevant. Some of the columns in the table will contain useful information for other sections of the systematic review such as when evaluating the quality of the studies, so some review authors may choose to add a column to this table for each study's quality assessment score. The background diet and other lifestyle variables of participants have to be considered when assessing study quality and therefore, it might be useful to document this information in the table alongside the other key information. An example of a study that been tabulated according to these requirements in the guidance document provided in the FSANZ guidance document (Appendix 2, page 26).

Where possible, make sure that the actual numbers from the tables/figures in the studies are reported for the sections on study results, rather than using what the authors of the study have described in the text of the article. It might also be helpful to compile a summary table for study results, which may be useful when considering whether there is a consistent association between the food and the health effect.

It is also required by Schedule 6 to document any adverse effects caused by the food or property of the food. For example, if examining the effects of a food or property of a food on bowel function, at one end of the scale are constipation-like symptoms such as infrequent bowel movements and stools that are hard and difficult to pass, and on the other end are diarrhoea-like symptoms such as very frequent bowel movements and stools that are watery and liquid-like. Both ends of the scale are also associated with gastrointestinal discomfort. Even if there is an average improvement in bowel function for the intervention group, it is also important to identify whether any participants experience adverse effects such as diarrhoea or constipation.

7.2 Assess methodological quality and applicability of studies

In order to obtain reliable conclusions from the systematic review and to comply with Standard 1.2.7, it is essential that when interpreting the results, they are considered in light of the quality of the studies. Study quality relates to an assessment of whether a study was designed and implemented appropriately (Higgins *et al* 2011). The publication of a study in a peer-reviewed journal (even a notable journal) does not guarantee that it will be high quality. Under Schedule 6, components of study quality that **must** be considered in the systematic review are outlined as follows:

Required elements of a systematic review - Schedule 6—2 (e):

> An assessment of the quality of each included study based on consideration of, as a minimum:
>
> (a) a clearly stated hypothesis
> (b) minimisation of bias
> (c) adequate control for confounding
> (d) the study participants' background diets and other relevant lifestyle factors
> (e) study duration and follow-up adequate to demonstrate the health effect
> (f) the statistical power to test the hypothesis.

Not all of these components will be important for rating the overall quality of each type of study included in the systematic review. For instance, assessing whether there is a clearly stated hypothesis is more important when assessing the quality of cohort studies. Therefore, if the types of studies included in the systematic review are all randomised intervention studies, then this component must be considered but might not be important to the overall quality rating in comparison with other quality indicators such as adequate statistical power to test the hypothesis. It is suggested that the assessment of the quality components is tabulated for each study included in the systematic review. This will allow the reviewers to get a clearer idea of the overall quality of each study included in the systematic review. It is recommended that at least two reviewers carry out the quality ratings separately and any differences be resolved by discussion or by a third reviewer.

7.2.1 Assess whether there is a clearly stated hypothesis

Considering whether a study has a clearly stated hypothesis is important for understanding whether the study was designed to investigate the food-health relationship that is the subject of the systematic review or whether the study set out with different aims and objectives. This is particularly relevant for cohort studies that have a large number of dietary variables that can be related to a number of different health outcomes (Smith and Ebrahim 2002). This should also be considered in an intervention study where the outcome might not be a primary or secondary objective mentioned in the original study protocol. This is important to consider if the health effect is only shown in a subgroup of the study population where that particular subgroup was not part of the original study hypothesis.

Sometimes there is no explicit hypothesis stated (that is, the predicted direction of the food-health relationship) in the study report but this could be found in a study protocol if one is available in a clinical trials registry. It is important to note however, that sometimes the primary outcome reported in the trial registers differ to that reported in the paper. Therefore, it

may be best to access the full study protocol in order to judge which information is correct. Otherwise, aspects of the hypothesis can be identified from key information in the study objectives or aims.

It is also important to assess whether the study reported all outcomes that were proposed to be measured. If there is incomplete or selective reporting of certain outcomes then this might bias the results of the study.

7.2.2 Assessing bias in intervention studies

The notion of bias is defined in the FSANZ guidance document as the *"Systematic deviation of a measurement from the 'true' value leading to either an over- or underestimation of the treatment effect"*. Bias can result from a number of different sources;

Selection bias may occur in the way that participants are allocated to groups if the randomisation procedure is inadequate or if there is foreknowledge of the allocation sequence which could lead to the allocation of participants to the groups being changed.

Performance bias may occur because of systematic differences in the behaviour of researchers or participants that may influence the health outcome due to knowledge of the intervention received.

Detection bias refers to systematic differences that may arise because of the way a health effect is assessed by the research personnel if they are not blinded to the intervention groups. **Attrition bias** refers to systematic differences between the groups in the number of participants who have withdrawn or been lost to follow-up from the study.

Reporting bias is the selective reporting of favourable outcomes within a study.

Other biases that may occur in a randomised trial are specific to particular study designs and may include carry-over effects of the intervention in cross-over studies without an adequate washout period (Higgins *et al* 2011).

7.2.2.1 Selection, Performance and Detection Bias

It is important that where a study compares an intervention (food or property of the food) with a control or a comparator group, that the participants in these groups are as alike as possible in all aspects that matter for the health effect, before the study begins. That is, they are matched for confounding variables such as age, sex, ethnicity, body mass index and education, but also other unmeasured confounding variables (Chalmers *et al* 2014). In studies with an appropriate sample size, allocating participants to groups by random lot will ensure appropriate matching of participants between the groups so that selection bias is minimised (Haynes *et al* 2012). It is important that the method of randomisation is appropriate; methods such as alternate allocation or allocation by date of birth are not appropriate because they are not random (Higgins *et al* 2011). Knowledge of the upcoming allocation sequence by those in charge of enrolment and allocation to the groups may lead to the selective assignment of participants to the groups if the sequence of groups to be allocated, or the sequence of participants who are enrolled, are changed (Pildal *et al* 2005). Therefore, it is important that the random allocation sequence is kept concealed (concealment of allocation), although methods such as sealed

envelopes can also be subverted easily. Empirical evidence shows that when concealment of allocation is absent or not known, this can lead to an exaggeration of the intervention effects especially where the outcome used is subjective e.g. self-reported ratings of mood or pain (Wood *et al* 2008). Many studies are described as being 'randomised' but if they lack an adequate method of randomisation or if the method of randomisation is not described, then they might not be randomised trials and could instead be "controlled clinical trials". These studies will still provide important information to support a food-health relationship but the results will have to be interpreted in the light of bias introduced by possible differences in confounding variables between the groups. Sometimes where the method of randomisation is not described it may be useful to look at the study protocol to see whether there is more information reported about the method of randomisation.

If a study has reported adequate blinding (or masking) of participants and research personnel, this means that no-one is aware of who is receiving the intervention or the control/comparator for the duration of the study. This is usually referred to as a "double-blind" design. It is also important to ascertain whether those measuring the health effect are also blinded to the groups that the participants were assigned. Evidence from a combined analysis of a number of meta-analyses shows that the absence of blinding will lead to an over-estimation of the effect of the intervention in studies where the outcome used is subjective e.g. self-reported ratings of mood or pain (Wood *et al* 2008). These findings may have particular relevance to a systematic review of a food-health relationship because in many dietary studies, participants cannot be blinded to the intervention they are receiving due to the intervention being a food with no suitable placebo. Therefore, it will be important to carefully consider the degree of impact that type of bias may have introduced in these studies.

7.2.2.2 Attrition Bias

Almost all randomised trials will have some missing values for outcomes. Attrition bias may occur when there are systematic differences between groups in the number of participants who drop-out or withdraw from the study between the intervention and control group because the analyses will deviate from the intended randomised comparison (Higgins *et al* 2011; Pocock and Abdalla 1998). Ideally, participants' last known data are carried forward and used as the final values in the analysis – this is a true *intention-to-treat analysis*. The degree of bias caused by attrition will depend upon the number of participants with missing data on the health outcome, and the extent to which there is an imbalance in missing data between the groups. Serious bias will occur if all of the participants who are deemed to be non-compliant with the intervention are excluded but subjects who are non-compliant with the control treatment are not excluded from the analysis. This creates imbalance between the groups in the study, especially where the reasons for not complying are related to a lack of effect or adverse effects of the intervention. If a study only reports this type of result then it should be excluded from the review, unless the intention-to-treat analysis can be reconstructed. Depending on the main reasons for non-compliance, this could lead to the effects of the intervention being over or underestimated (Pocock and Abdalla 1998).

7.2.2.3 Reporting Bias

Reporting bias refers to the selective or incomplete reporting of outcomes in a study. Examples of reporting bias can include not presenting the full amount of data for the main outcome (e.g. where the outcome is measured and analysed several times throughout the study) or not reporting the actual results for an outcome (e.g. only reporting *p* values for the

results or simply stating that results were not statistically significant without giving the results or the exact *p* value). Evidence shows health outcomes that are statistically significant have a greater likelihood of being fully reported compared to health outcomes that are not statistically significant (Chan *et al* 2004). Therefore, if there is evidence of reporting bias, this might exaggerate the effect of the intervention.

7.2.3 Assessing bias in observational studies

Considering whether there has been adequate follow-up is important when trying to gauge the impact of possible bias in a cohort study. Participants who are lost to follow-up (e.g. those who discontinue participation in the cohort study, move without notifying the study co-ordinators, or die) are likely to be different from participants who continue their participation in the cohort study in terms of their dietary exposure and other important characteristics such as education and smoking status.

In terms of assessing bias, it is also important to consider whether participants have been correctly classified into the appropriate category of dietary exposure. This might depend upon the instrument used to measure dietary intake, the number of times that diet was assessed (especially if the period of follow-up is a number of years), and whether there is also an objective measure of dietary intake (i.e. nutritional biomarkers).

7.2.4 Assessing whether there has been adequate control for confounding

In the FSANZ guidance document (page 23) confounding is defined as *"The measure of the treatment effect is distorted because of the differences in variables between the treatment and the control group that are also related to the outcome."* Confounding can also be referred to as a confusion of effects (Vandenbroucke *et al* 2007). Adequate control for confounding is an important aspect of study quality and it is essential to ensure that the effect on health is due to the food or property of the food of interest and not due to some other factor. In randomised trials that have achieved adequate randomisation (i.e., an appropriate method of random allocation to groups has been well described) then control for confounding would be considered adequate (Higgins *et al* 2011). It might also be useful to consider whether the baseline characteristics between groups are similar in terms of the measured confounding variables. For studies where the method of randomisation used is not clear, then it is possible that control of confounding is inadequate. Some randomised trials may have performed an analysis that includes an adjustment for baseline values of certain variables (e.g. baseline measures of the health effect, age). If this is done, then control for confounding would be considered adequate (Kirkwood and Sterne 2003). However, when studies have small sample sizes, there may be uneven distribution of values, even if excellent randomisation has been done. In this case, confounders should not be adjusted for in the analysis.

Often observational studies such as cohort studies are used to investigate a number of different health outcomes and for each separate analysis, different inclusion and exclusion criteria are applied to generate the final study population. One of the important factors to consider when assessing bias in a cohort study is the extent to which the participants who may have a condition that affects both the health effect and exposure (reporting of diet) have been excluded from the analyses. For example, if a cohort study assesses the association between intake of omega-3 fatty acids at baseline and attention/concentration in children measured several years later, then it would be important that children with attentional disorders (or very low scores of attention/concentration) are excluded from the analysis. This is because

inclusion of these children may bias the results towards a more favourable effect because children with very low attention may have a much lower intake of all foods, including omega-3 fatty acids.

There should also be an assessment of whether all important confounding variables have been included in the statistical model as covariates in an observational analysis. It is also important to judge whether after adjusting for a confounding variable, all the effects of that variable have been removed. For example, where the continuous variable body mass index (BMI) might confound the relationship between a food or property of a food and an outcome, then careful consideration should be given to the way that BMI is included in the statistical model. Adjusting for BMI as a categorical variable (two or more categories) might not capture all the information necessary that is required to achieve sufficient statistical control and it is possible that it will result in inadequate control of confounding, which is often referred to as residual confounding (Royston *et al* 2006). Sometimes even very sophisticated statistical analyses cannot disentangle the effects of confounding variables from that of a dietary variable because they are often inextricably linked, and the results of observational studies should be interpreted in light of this.

Whether a study has "over adjusted" for confounding variables is something that ought to be assessed, as this may underestimate the effect of the food or the property of the food on health. This is important in the case where variables that may mediate the relationship between the food or the property of the food and the health effect, are included as covariates. For example, where intervention studies examining the effect of an energy-reduced food on body weight have adjusted for total energy intake, this would be considered "over adjusting" as the effect of an energy-reduced food would be through an effect on total energy intake.

7.2.5 Assessing study participants' background diets and lifestyle factors

When considering the background diets and other lifestyle factors of the participants in each study included in the systematic review, it is important to evaluate whether the effect of the food (or property) is observed in all study populations or in a subgroup of the studies. For example, if the health effect of fish oil supplements is only observed amongst participants who consumed a background diet high in fish but not in participants with a lower background intake of fish, this might suggest there is a minimum level of exposure to fish oils that is required to produce the health effect. If the health effect of fish oils is only present in studies whose participants had a low background intake of fish, this might suggest a threshold effect.

Another reason for considering the participants' background diet and other lifestyle factors relates to the external validity of the evidence. External validity refers to the generalisability of the results; in other words, are the results applicable to another population or the target population? If the participants selected for the study differ from that of the intended target population, this might limit whether the results are applicable. This is especially true if the study population is generated from a workplace where people who are working will tend to be healthier than those not working (Delgado-Rodríguez and Llorca 2004). The generalisability of the findings from the studies will also be considered in the context of Australian and New Zealand populations, later on in the systematic review. Notwithstanding, the importance of considering the external validity of a study, it is noteworthy that the internal validity of a study is more important (e.g. bias), as it is not appropriate to generalise invalid results.

7.2.6 Assessing adequacy of study duration and follow-up

The exact duration of a study that is expected to demonstrate the effect of a food or a property of a food on a health effect will depend upon the proposed effect on health. If the health effect relates to postprandial health effects then for example, studies with duration of half to two hours would be adequate to demonstrate the effects on blood glucose, whereas six to eight hours would be considered adequate to demonstrate an effect on blood triglycerides. For other health effects it might be useful to factor in the amount of time that these processes usually take. For example, the health effect 'colon transit time' or 'whole gut transit time' can take 72 hours or more (depending on the individual), so a study with a duration of three weeks would be sufficient to see an effect on these outcomes but a study that ran for three days would not. To demonstrate an effect on most analytes in the blood, usually a study with a duration of a month would be sufficient to see an effect. It is likely that the duration of studies to see relevant changes in other health effects such as cognition and growth might be longer (months to years).

Where cohort studies are used as evidence to support a food-health relationship, the time between dietary intake and the health effect being assessed may be quite long given the logistics of collecting information from a large number of participants. In the case of a long period of follow-up, the reviewers should decide whether this time frame is appropriate for the proposed effect on health.

7.2.7 Assessing whether there is sufficient statistical power to test the hypothesis

Depending on the area of research, the majority of randomised trials would not report a power calculation in the methods section or this calculation might be reported in the study protocol which might be available in clinical trial registries, if the trial is recent and registered. If a study does not contain a sufficient number of participants or participants who experience the outcome (where the outcome is binary – yes/no), there will be greater uncertainty (i.e. wider confidence intervals) about the true effect of the intervention. It is also important to assess the statistical power to test the hypothesis where small studies report a large effect size that is statistically significant. In this case, it is possible that a statistically significant result from a small, underpowered study does not reflect the true underlying effect (Button *et al* 2013), especially if the results are not consistent with other similar studies with larger sample sizes.

7.3 Quality appraisal tools

The guidance document by FSANZ lists six different tools that can be used to assess certain aspects of study quality such as the risk of bias (reviewers are not limited to these six tools). While several of the tools are referred to as "quality assessment tools", none of these tools assess all components of study quality that are laid out in Schedule 6 of Standard 1.2.7. It is important that all six aspects of the studies included in the systematic review are assessed in order to meet these requirements.

Some of these quality appraisal tools use a scale or assign scores to individual items based on certain characteristics of a study to give an overall score or rating that can be used to categorise the study according to its quality. Other tools do not assign a score but instead evaluate the risk of bias within a study for each of the outcomes assessed. Each tool will give a slightly different grading simply because different questions are asked or different aspects of a study are assessed and given different weighting. It is up to those conducting the systematic review to decide which tool, or combinations of tools to use – some are more straightforward to complete than others. As mentioned in the FSANZ guidance document, it is recommended that at least two reviewers complete the chosen quality appraisal tool and compare their ratings.

7.3.1 Quality appraisal tool for intervention studies from Health Canada

The quality appraisal tool for intervention studies from Health Canada has been designed specifically for the assessment of the quality of studies included in a systematic review for the substantiation of health claims in Canada. There are 15 aspects of the study that the reviewer must assess. A score of one is assigned to each item if it is present in the study and a score of zero if it is not present (or not reported), and if the score totals eight or more, then the study is considered "higher quality" but if the score is less than eight, then the study is considered "lower quality" (note that a "higher quality" study might not necessarily be a "high quality" study according to Standard 1.2.7). However, the Health Canada quality appraisal tool covers two aspects of study quality (bias and confounding) that are outlined in Schedule 6 of Standard 1.2.7. Therefore, if the study is categorised as being "higher quality" then this applies to the bias and the confounding aspects of study quality but the "higher quality" might not apply to the other aspects of study quality that are part of Schedule 6.

7.3.2 Cochrane Collaboration's tool for assessing the risk of bias

Another tool – the Cochrane Collaboration's tool for assessing risk of bias – has been designed specifically for assessing the risk of bias in randomised studies included in a Cochrane systematic review (Higgins *et al* 2011). Additional guidance on how to complete this risk of bias table is also provided by the Cochrane Bias Methods Group. This tool does not assign a score to any of the items included but instead evaluates the risk of bias (low, unclear or high) for each main outcome (or class of outcomes) measured in an intervention study. Note that a low risk of bias will indicate a better rating for this aspect of study quality.

For **each of the key domains** included in this table, a judgement is made about the risk of bias that it might pose:

Low risk of bias – bias, if present, is unlikely to alter the results seriously
Unclear risk of bias – a risk of bias that raises some doubt about the results
High risk of bias – bias may alter the results seriously

Once a judgement has been made about the risk of bias for each of the components included in this tool, the overall risk of bias for **each study** can be summarised as follows:

Low risk of bias – low risk of bias for all key domains
Unclear risk of bias – low or unclear risk of bias for all key domains
High risk of bias – high risk of bias for one or more key domains

After each study has been summarised the following judgment can be made across **all studies**:

Low risk of bias – most of the information is from studies with a low risk of bias;
Unclear risk of bias – most of the information is from studies at a low or unclear risk of bias.
High risk of bias – the proportion of information from trials at high risk of bias is sufficient to affect the interpretation of results

Other study quality appraisal tools that are listed in the FSANZ guidance document such as the checklist to appraise the quality of interventions that is recommended by the National Health and Medical Research Council (NHMRC) in Australia contains items that are similar to the Cochrane Collaboration tool. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) tool uses a slight adaptation of the Cochrane Collaboration tool to assess bias in randomised controlled trials when assessing the quality of the evidence.

The different quality appraisal tools will give slightly different overall ratings. While the quality appraisal tool for intervention studies from Health Canada is reasonably easy to complete and allows for a simple deduction of whether the study is of low or high quality, this tool was designed specifically for assessing the quality of studies for a submission to Health Canada for a food health claim Canada. Although there are similarities between Canada and New Zealand for many aspects of the scientific substantiation of health claims, the components of study quality that must be taken into consideration are different. Under Standard 1.2.7 there are additional aspects of study quality (study hypothesis, participants' background diet and lifestyle factors, study duration and follow-up, and statistical power) that need to be considered and factored into the overall rating of study quality when self-substantiating a food-health relationship (see Sections 6.2.2 and 6.2.6 to 6.2.8 for more information). The Cochrane Collaboration's tool assesses the risk of bias but other components of study quality (study hypothesis, confounding, participants' background diet and lifestyle factors, study duration and follow-up, and statistical power) will still need to be assessed (see Sections 7.2.1 and 7.2.4 to 7.2.7 for more information) and factored into the overall rating of study quality.

7.4 Evidence from observational studies to support a food-health relationship

The majority of high quality evidence to support a food-health relationship will come from intervention studies (preferably randomised trials) but there may be cases where evidence from observational studies such as cohort studies provide evidence to support the food-health relationship such as is the case for several of the pre-approved food-health relationships listed in Standard 1.2.7. A common difficulty in using evidence from observational studies is deciding whether the association is due to the food or property of the food, or some other confounding variable (that may or may not be measured in the study). Sometimes even very sophisticated statistical analyses cannot disentangle the effects of confounding variables from that of a specific dietary variable because they are often inextricably linked. The results of observational studies must therefore be interpreted in light of this potential confounding.

7.4.1 Health Canada's quality appraisal tool for observational studies

The quality appraisal tool for observational studies from Health Canada has been designed specifically for the assessment of the quality of observational studies included in a systematic review for the substantiation of health claims in Canada. There are 12 questions relating to aspects of the study that the reviewer must assess. A score of one is assigned to each item if it is present in the study and a score of zero if it is not present (or not reported). If the score totals seven or more, then the study is considered "higher quality" but if the score is less than seven, then the study is considered "lower quality". However this tool for observational studies covers only two of the required aspects of study quality (bias and confounding) that are outlined in Schedule 6 of Standard 1.2.7.

7.5 Random measurement error

One of the main points of having a suitable control group within an intervention trial is to test what would happen over the duration of the study if they received no intervention or received 'usual diet'. This is important in order to rule out 'regression to the mean' phenomena where measurements of the health outcome can show an average change in the absence of any intervention. This is because many health effects are measured with random error that is due to technical measurement error and real within person fluctuations (Whitlock *et al* 2001; Vandenbroucke *et al* 2007). Take for example, a group of participants who are selected based on having extreme values of the health effect of interest (e.g. those with a low number of weekly bowel movements) who may be given an intervention to increase the number of weekly bowel movements. When these participants have the frequency of their bowel movements measured again the mean of group the will be closer to the mean of the wider population – that is, it has increased (conversely, participants with a greater number of weekly bowel movements would show a decrease towards the mean after re-measurement). This should not be interpreted as showing an effect of the intervention because even if participants are not treated the mean number of weekly bowel movements will go up, owing to regression towards the mean. Therefore, having a suitable control group that have their bowel movements measured over the same period of time will allow for any effect of regression to the mean to be factored into the analysis.

Some of the quality appraisal tools ask whether the exposure variable was assessed more than once in the observational studies. This attempts to account for the effect of random error

(sometimes referred to as measurement error). Dietary assessment methods such as food frequency questionnaires, 24 hour recalls and food diaries, and nutritional biomarkers (nutrients measured in the blood and other biological tissues) all measure diet with a certain degree of error (that is, the measured intake will deviate from the true intake). Generally the exposure of interest in cohort studies would be "usual" dietary intake as this will be related to the health outcome sometime after diet is assessed at baseline. Measuring dietary intake or nutritional status at only one point in time might not capture usual intake adequately, and this will tend to underestimate the strength of the association between the food or property of the food and the health effect. This is mainly because the extreme categories include more participants than they should. For example, the lowest category of dietary fibre intake will have more participants whose measured intake of dietary fibre is lower than their usual dietary fibre intake, whereas the top category of dietary fibre intake will include more participants with higher measured dietary fibre intake than their usual intake. Cohort studies that have repeat dietary intake measures (or nutritional biomarkers) on a subgroup of the study population can correct for some of this measurement error and the resulting association with the health effect will tend to be stronger (MacMahon *et al* 1990).

7.6 Statistical analyses in intervention studies

Some of the quality appraisal tools ask whether the statistical analysis was appropriate for the study design or whether a between-group statistical analysis of the health effect was conducted. In order to assess the effect of the intervention in a trial, it is important that the statistical analysis compares the size of the health effect between the groups (that is, the intervention vs. control) by using a two-sample t-test (or equivalent). Testing whether there has been a statistically significant change in the health effect within each group (i.e. comparison of end of study values with baseline values in each group) and then declaring that there is a significant effect if only one group has a significant change, is a biased method of statistical analysis of a intervention study and will greatly inflate the level of type I error (i.e. rejecting the null hypothesis when it is true) (Bland and Altman 2011). To properly test the effect of an intervention in a group of people, there needs to be a control group.

When considering the results of a study, in addition to considering whether the difference between the intervention and control group is statistically significant (usually $p < 0.05$), it is also good to judge the precision of the effect of the intervention by considering the width of the confidence intervals around the effect size estimate. The 95% confidence intervals give an indication of the range in which the true effect of the intervention could possibly lie (Guyatt *et al* 2011). Because of the way that the confidence intervals are calculated, studies that have a greater sample size or have a larger number of participants who experience the outcome (where the outcome is binary – yes/no) will have narrower, more precise confidence intervals. The precision of each of the study results is an important aspect to factor in when judging whether there is causal relationship between the food or property of the food. It is recommended that someone with expertise in statistical analysis is also consulted to assess the appropriateness of the analysis.

7.7 Valid measures of the health outcome

Some of the quality appraisal tools ask about the validity and the reliability of the data collection tools. This can apply to both the methods used to assess the exposure e.g. dietary assessment techniques and those used to measure the health outcome. In the case of the effect on health, it is important that the methods used to measure the health outcomes are valid (i.e., they measure what they intend to measure) and reliable (i.e., repeated measures give a similar value). If a questionnaire is used to assess a component of health, it is usually stated whether it has been validated in the paper.

8 Make an overall decision

Required elements of a systematic review - Schedule 6—2 (f):

> An assessment of the results of the studies as a group by considering whether:
>
> (a) there is a consistent association between the food or property of food and the health effect across all high quality studies
> (b) there is a causal association between the consumption of the food or property of food and the health effect that is independent of other factors (with most weight given to well-designed experimental studies in humans)
> (c) the proposed relationship between the food or property of food and the health effect is biologically plausible
> (d) the amount of the food or property of food to achieve the health effect can be consumed as part of a normal diet of the Australian and New Zealand populations.

8.1 Consistent association across all high quality studies

After all the study results have been tabulated and the quality of the studies has been assessed, it is important to consider the totality of the evidence for an effect of the food or the property of the food on the health outcome, across all high quality studies. Evaluating the consistency of an association can be quite difficult, especially for a health effect that is not measured using the same methods across all studies.

It is not necessary to conduct a meta-analysis (statistical method that combines all of the data on the health outcome across the studies) to meet the requirements of Schedule 6 in Standard 1.2.7, but if it is possible to do so, then this is a more straightforward way to assess the consistency of the association. There are a number of statistical packages that allow for a meta-analysis to be done.

It is recommended that a statistician is consulted before combining all the results into a meta-analysis as there are a number of assumptions underlying the different types of ways to combine the data which can affect the overall results.

The guidance document by FSANZ (page 14) contains a link to a tool from Health Canada to combine study results in a way that does not involve a meta-analysis. This tool only considers whether the results of (high quality) studies are statistically significant and does not consider the size of the study or more importantly the precision of the study results in rating the consistency of study results. It is suggested that studies which have a larger sample size and give a more precise estimate of the true effect size are given a greater weighting to the overall

assessment of consistency. For example, if the results of three smaller high quality studies (e.g. fewer than 50 participants in each study) show that the interventions have a positive effect on the health outcome that are statistically significant with wide confidence intervals, but the results of two larger high quality studies (e.g. more than 200 participants in each study) show that the effect of the interventions are not statistically significant with narrow confidence intervals, then the overall consistent association would be that there is no effect on the health outcome.

Usually the effect of the food or property of food and health outcome would have to be demonstrated in more than one high quality study. This is because there are many cases where an initial positive statistically significant finding has been published and disseminated but this relationship has not been supported by subsequent studies, which are usually larger and of higher quality (Ioannidis 2005). As stated in the guidance document by FSANZ, it is difficult to give an estimate of the number of high quality studies needed to show a consistent effect of the food-health relationship, as amongst other factors, it will depend on the size, and the precision of the studies' results.

8.2 Causal association between the food and the health effect

The guidance document by FSANZ (pages 14 and 15) provides useful information of some of the Hill's criteria for causation that can be used in assessing whether the association between a food or the property of a food and a health effect is causal. It is important to note that these causal criteria were developed in light of research that was being undertaken to determine whether specific occupational hazards were causing diseases such as cancer from observational studies; however, these criteria are still pertinent for establishing a causal relation between a food or the property of a food and a health effect.

8.2.1 Consistency of results

Please refer to the description of a 'Consistent association across all high quality studies' (Section 8.1) for information on assessing the consistency of results.

8.2.2 Strength of the association

The greater the strength of an association between a food or a property of a food on a health effect, the greater the likelihood of the association being causal; however, this does not mean that smaller effects are not likely to be causal (Hill 1965). It is difficult to provide a threshold that will constitute what is considered to be a 'strong' association; however, if results show an effect on the health outcome that in an order of 40-50% then this would probably be considered strong (Potischman and Weed 1999).

Often smaller studies that have been published will have produced a greater effect on the health outcome than larger studies simply because smaller studies usually have wider confidence intervals and therefore, a much larger effect size is required for the results to be statistically significant (Button *et al* 2013). Therefore, it is important to take the size of the studies and the width of the confidence intervals into consideration when assessing the strength of the association.

Where studies have not measured a health effect using the same methods it might be difficult to assess the overall strength of the association (size of the effect) given that different scales or units of measurement will give a different magnitude of effect. In this case, it might be more meaningful to standardise the average effect size in each study so that the size of the effect across the studies is comparable.

8.2.3 Dose-response relationship

This refers to a greater effect of the food or the property of food on the health outcome with increasing amounts of the food consumed. The presence of a dose-response relationship may strengthen the likelihood that there is a causal association but absence of a dose-response relationship does not mean that the association is not causal. It might not be possible to assess a dose-response in intervention studies where similar amounts of the food are given across the studies. It is also possible that the food or the property of the food has a threshold effect, beyond which the effect on the health outcome is minimal. For assessing a dose-response relationship in observational studies, it is also important to take account the misclassification that may occur when participants are separated into increasing categories of intake for a food or property of a food, which may weaken or obscure any dose-response relationship (Potischman and Weed 1999 see section 7.5 'Random measurement error' for more information).

8.2.4 Temporality

It is important for determining whether there is a causal association between a food or a property of a food and an effect on a health outcome, that the exposure (food) is measured before the effect on health (Potischman and Weed 1999). This will be evident in randomised trials where the food or property of the food is given to the participant and the effect on the health outcome occurs after ingestion. The measurement of diet in cohort studies almost always takes place before the effect on health is measured. For retrospective case-control and cross-sectional studies, it is often difficult to establish whether the food or property of the food causes the effect on health or altered intake of that food is a result of an effect on health (Hill 1965) and therefore evidence from these types of studies will not be able to substantiate a health claim.

8.3 Biologically plausible association between the food and the health effect

Animal, *in vitro* and human metabolic studies can all provide information to support a biologically plausible effect of a food-health relationship. These studies should provide evidence for an effect of the food or the property of the food on human health by providing insights into possible mechanisms through which the food exerts the effect. The extent to which the food or the property of the food demonstrates biological plausibility in these studies will obviously depend on where the science currently stands (Hill 1965).

If the claimed effect on health by the food or property of the food is deemed to be causal then demonstrating a biologically plausible pathway by which the food or property of the food could have an effect on health will strengthen the likelihood of causality. However, if there is little or no evidence of a causal relationship between the food or the property of the food on the health effect from high quality intervention studies in humans, then studies that demonstrate biological plausibility will do nothing to strengthen the relationship.

Required elements of a systematic review - Schedule 6—2 (g):

> A conclusion based on the results of the studies that includes –
>
> (a) whether a causal relationship has been established between the food or property of food and the health effect based on the totality and weight of evidence; and
> (b) where there is a causal relationship between the food or property of food and the health effect:
> (i) the amount of the food or property of food required to achieve the health effect
> (ii) whether the amount of the food or property of food to achieve the health effect is likely to be consumed in the diet of the Australian and New Zealand populations or by the target population group, where relevant.

8.4 Whether a causal relationship has been established

It is stated in the guidance document by FSANZ that *"One way of thinking about causality might be to consider whether it is likely or not that another large, well-conducted study would have such different results from the available studies that the conclusion from the systematic review would be altered importantly."* This is demonstrated in an analysis by Egger *et al* (1997) where the results from several systematic reviews show a statistically significant effect on health that are not demonstrated in later large randomised controlled trials. On this basis, the conclusion of these systematic reviews regarding the overall effect on health would be altered considerably.

8.5 The amount of the food or property of the food to achieve the health effect

There is a detailed description of factors that need to be considered when deciding whether the amount of the food or the property of the food to achieve the health effect could be consumed as part of a normal diet in Australian and New Zealand populations in the FSANZ guidance document. For instance, if the health claim was for a daily serving of fish, it might be unreasonable to expect that this level of fish could be consumed by the general New Zealand population given that more than half of the population consume fish less than once a week (University of Otago and Ministry of Health, 2011).

Required elements of a systematic review - Schedule 6—2 (h):

> An existing systematic review may be used if it is updated to include –
>
> (a) the required elements 1 to 6 above for any relevant scientific data not included in the existing systematic review
> (b) the required element 7 above incorporating the new relevant scientific data with the conclusions of the existing systematic review.

Similar guidance would apply to all the relevant sections in the case where an existing systematic review is being updated to substantiate a food-health relationship.

9 Developing a health claim from a substantiated food-health relationship

The systematic review is the process of establishing whether there is a causal relationship between the food or the property of the food and the effect on health. Once the food-health relationship has been established, the wording for the general level health claim can be developed. For the pre-approved food-health relationships, none of the wording in Standard 1.2.7 is taken to be prescribed when developing the health claim and any statement or information may be modified if the modification does not alter or contradict the meaning of the pre-approved relationship.

It is important that the wording of the health claim for an established food-health relationship is comprised of the components that make up the relationship – the food or the property of the food and the health effect. When making a health claim, food businesses will want to use a language that is readily understood and enticing to their consumers but the wording of the health claim must be a true and accurate representation of the substantiated food-health relationship. It is important that the wording of the health claim does not convey an effect on health that is too broad or take the health effect to a level beyond what the scientific evidence has demonstrated. For example, if a health claim for the food-health relationship "[name of the nutrient] is necessary for normal immune system function" read "contains [name of the property of the food] for a healthy immune system", this would not be considered to alter the health effect. Conversely, "contains [name of the property of the food] to boost your immune system" might be considered as amplifying the stated food-health relationship because "boosting your immune system" is a step beyond maintaining a normal (healthy) immune system, which is the effect on the immune system that has been scientifically substantiated.

Likewise,, a substantiated food health relationship of “contains x for regular laxation” is not equivalent to a claim of “contains x for good digestive health”, since the former is only one aspect of the latter.

As all health claims must be accompanied by a dietary context statement, for a self-substantiated food-health relationship, the dietary context statement must also be consistent with the conclusions of the systematic review.

10 References

Altman and Bland (1998). Generalisation and extrapolation. BMJ 17:409–10.

Bland JM, Altman DG (2011). Comparisons against baseline within randomised groups are often used and can be highly misleading. Trials 12:264.

Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience 14(5):365-76.

Chalmers, Iain (2003). Trying to do more good than harm in policy and practice: the role of rigorous, transparent, up-to-date evaluations. The Annals of the American Academy of Political and Social Science 589.1 (2003): 22-40.

Chalmers I**,** Tröhler U, Clarke M (editors). JLL Explanatory Essay 4.1. Differences between the people compared. The James Lind Library: www.jameslindlibrary.org (Accessed 16 July 2014).

Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG (2004). Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 291(20):2457-65.

Delgado-Rodríguez M, Llorca J (2004). Bias. Journal of Epidemiology and Community Health 58(8):635-41.

Egger M, Davey Smith G, Schneider M, Minder C (1997). Bias in meta-analysis detected by a simple, graphical test. BMJ 315(7109):629-34.

Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux PJ, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW Jr, Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J, Schünemann HJ (2011). GRADE guidelines 6. Rating the quality of evidence--imprecision. Journal of Clinical Epidemiology 64(12):1283-93.

Haynes L, Service O, Goldacre B, Torgerson D: Behavioural Insights Team (2012). Test, learn, adapt: Developing public policy with randomised controlled trials. Cabinet Office

Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savovic J,
Schulz KF, Weeks L, Sterne JA; Cochrane Bias Methods Group; Cochrane Statistical Methods Group (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 343:d5928.

Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2 [updated September 2009]. The Cochrane Collaboration, 2009. Available from www.cochrane-handbook.org.

Hill AB (1965). The environment and disease: association or causation? Proceedings of the Royal Society of Medicine 58:295-300.

Ioannidis JP (2005). Why most published research findings are false. PLoS Medicine 2(8):e124.

Kirkwood BR, Sterne JAC (2003). Essential Medical Statistics. Blackwell Publishing, Oxford, UK.

MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J, Abbott R, Godwin J, Dyer A, Stamler J (1990). Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. Lancet 335(8692):765-74.

Pildal J, Chan AW, Hróbjartsson A, Forfang E, Altman DG, Gøtzsche PC (2005). Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. BMJ 330(7499):1049.

Pocock SJ, Abdalla M (1998). The hope and the hazards of using compliance data in randomized controlled trials. Statistics in Medicine 17(3):303-17.

Potischman N, Weed DL (1999). Causal criteria in nutritional epidemiology. American Journal of Clinical Nutrition 69(6):1309S-1314S.

Royston P, Altman DG, Sauerbrei W (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. Statistics in Medicine 25(1):127-41.

Smith GD, Ebrahim S (2002). Data dredging, bias, or confounding. BMJ 21;325(7378):1437-8.

Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ (2000). Publication and related biases. Health Technology Assessessment 4(10):1-115.

University of Otago and Ministry of Health (2011). A Focus on Nutrition: Key findings of the 2008/09 New Zealand Adult Nutrition Survey. Wellington: Ministry of Health.

Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M; STROBE Initiative (2007). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Medicine 4(10):e297.

Whitlock G, Clark T, Vander Hoorn S, Rodgers A, Jackson R, Norton R, MacMahon S (2001). Random errors in the measurement of 10 cardiovascular risk factors. European Journal of Epidemiology 17(10):907-9.

Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 336(7644):601-5.