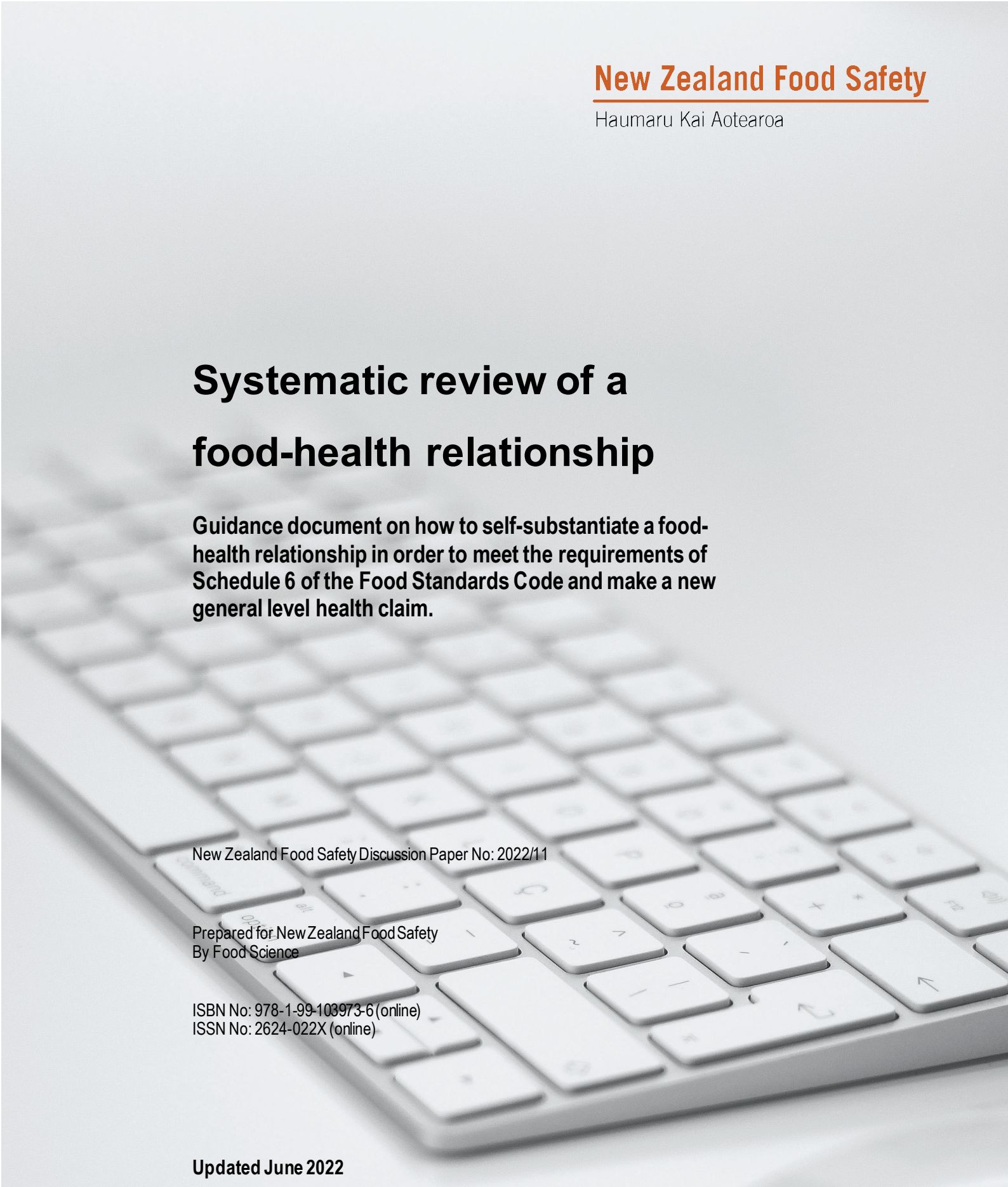
New Zealand Food Safety
Haumaru Kai Aotearoa

Systematic review of a food-health relationship

Guidance document on how to self-substantiate a food-health relationship in order to meet the requirements of Schedule 6 of the Food Standards Code and make a new general level health claim.



New Zealand Food Safety Discussion Paper No: 2022/11

Prepared for New Zealand Food Safety
By Food Science

ISBN No: 978-1-99-103973-6 (online)
ISSN No: 2624-022X (online)

Updated June 2022

About this document

This document is intended to provide food businesses and researchers 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 the Australian and New Zealand Food Standards Code: Standard 1.2.7 (Nutrition, health and related claims) and Schedule 6 (Required elements of a systematic review). While guidance material is not legally enforceable, stakeholders are encouraged to discuss significant departures from this best practice approach with NZFS 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, nor intended to be representative.

Contact details

New Zealand Food Safety
Food Science and Risk Assessment Directorate
PO Box 2526
Wellington 6140
Email: healthclaims@mpi.govt.nz

Disclaimer

This guidance does not constitute, and should not be regarded as, legal advice. While every effort has been made to ensure the information in this publication is accurate, the Ministry for Primary Industries does not accept any responsibility or liability for error of fact, omission, interpretation or opinion that may be present, nor for the consequences of any decisions based on this information.

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

Cover photo by Hugo Barbosa on Unsplash

Contents

1 Executive summary

Standard 1.2.7 of the Food Standards Code 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 demonstrating 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 and Schedule 6 of the Food Standards Code.

A systematic review is an overview of a specific research question that systematically identifies, selects, appraises, and synthesises all high-quality primary research evidence relevant to that question. It is essential that the systematic review includes the totality of the relevant scientific evidence, regardless of the study results. To meet the specific requirements of Schedule 6 of the Food Standards Code, the key steps of the systematic review are as follows:

- For the purpose of substantiating a food-health relationship, the food or property of the food and the health effect must be defined.
- Based on the food/property of the food and the health effect, the review question is developed
- The search strategy is formulated. This includes determining the inclusion/exclusion criteria for studies eligible for inclusion.
- Relevant human studies from multiple sources are identified, listed, and screened against predetermined inclusion/exclusion criteria.
- Evidence is evaluated: this involves constructing a table that includes information on key components of each of the studies included in the systematic review.
- Each eligible study is evaluated for study quality
    - Double-blind, randomised, placebo-controlled trials are preferred.
    - A study appraisal tool may partially assist in quality evaluation of studies. Further discussion of overall study quality is recommended, since a study quality appraisal tool will not assess all quality elements required in Schedule 6, and not all quality elements are of equal importance.
- The results from the high-quality studies are evaluated for consistency and causality.
- Biological plausibility of the food-health relationship is discussed.
- It is demonstrated that the food or property of the food can be realistically consumed as part of the normal diet of Australian and New Zealand populations, at the amount required to achieve the health effect.
- A conclusion is determined, based on whether the totality and weight of evidence establishes a causal food-health relationship.

Assessing overall study quality and deciding whether there is a causal relationship between a food (or a property of a food) and a health effect involves a certain degree of judgement. This guidance document is designed to provide industry with relevant insights into the processes used by New Zealand Food Safety (NZFS) to make such judgements.

2 Purpose of this guidance document

This document is intended 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. This document contains more detail, and provides more examples for each of the requirements of the systematic review outlined in Schedule 6 of the Food Standards Code. This 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'.

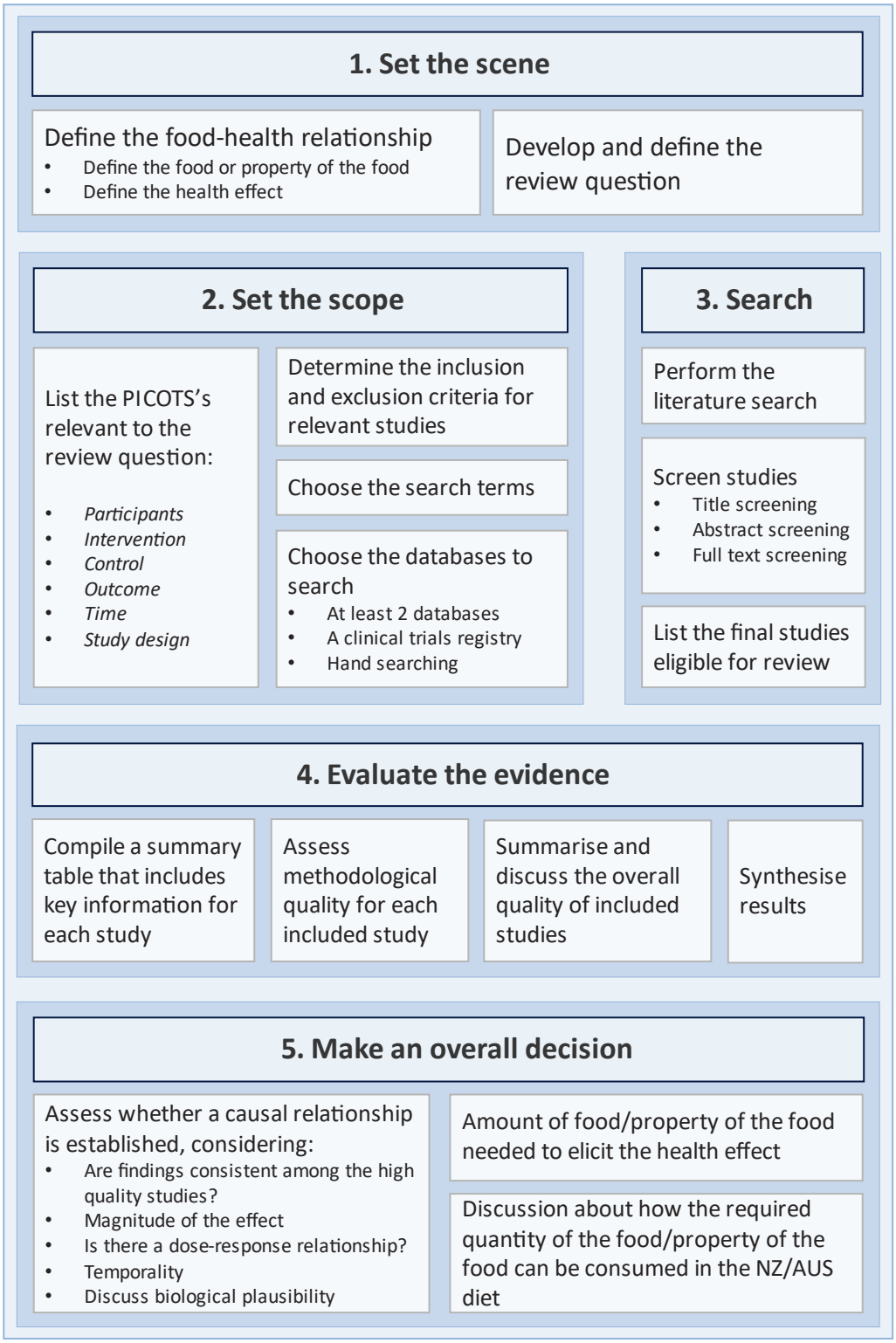
This document describes in detail each of the required elements of a systematic review, which are outlined in Schedule 6 of the Food Standards Code. The purpose of this document is not to describe how to do a systematic review, but rather to provide guidance on the requirements of Schedule 6. These requirements include elements that are not typically found in other published systematic reviews.

All components of a systematic review for the purpose of self-substantiating a food-health relationship are outlined in Figure 1).

If a New Zealand food business is considering completing a systematic review to self-substantiate a food-health relationship, New Zealand Food Safety (NZFS) would encourage the food business to make contact with NZFS, who will be able to provide regulatory and technical advice well in advance of the notification stage. NZFS will treat all information you provide with complete confidence and can provide detailed feedback and guidance specific to each food-health relationship. Please email NZFS at health.claims@mpi.govt.nz.

In addition, NZFS encourages food businesses to seek assistance from scientific experts who have skills and experience with compiling systematic reviews in nutrition, to write or update your systematic review. This is because, as you will see by this guidance document, to meet the requirements of Schedule 6 you need to have significant skills and experience in nutritional epidemiology, and access to multiple databases of scientific literature. Many such experts exist within Universities and specific research agencies, but there are also consultants who specialise in this type of work. NZFS can assist in explaining the regulatory requirements and commenting on draft research plans, protocols/search strategies and reviews in light of these requirements.

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



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. A systematic review includes the entirety of the relevant studies for a specific research question. This means that individual studies can be interpreted in the context of other similar studies, rather than considering single studies in isolation. To produce reliable results, the process of conducting a systematic review involves methodically locating, critically appraising, and synthesising the relevant 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. 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, resulting in an imbalance between the groups in the study. This imbalance will introduce bias if the reasons for not complying are related to a lack of effect or adverse effects of the intervention. To minimise the effect of bias, study quality assessments must be undertaken for each eligible study (see Section 7.2).

A systematic review conclusion is only reliable if the methods used to conduct the systematic review are appropriate. The methods include (but are not limited to) forming the systematic review question, designing the search strategy, determining the inclusion and exclusion criteria for individual studies, conducting study quality assessments, and interpreting the evidence. For example, a flawed search strategy may result in some eligible studies being missed, and this could introduce unwanted bias to the systematic review. (Chalmers 2003).

3.1 Guidance on New Zealand Food Safety's 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 the Food Standards Code.

Upon receiving information that a self-substantiated food-health relationship has been notified to FSANZ (or NZFS if it relates to a supplemented food) by a New Zealand food business, NZFS will request the dossier of evidence to support the food-health relationship and assess whether an evaluation is warranted.

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.

NZFS is the relevant authority in New Zealand and will ask to see the dossier of supporting evidence including the systematic review. NZFS then evaluate whether the dossier meets the requirements outlined in Schedule 6 and that food-health relationship is a reasonable conclusion of the systematic review. When NZFS asks to see the dossier, specific components of it are likely to be evaluated by NZFS, to ensure that the scientific evidence supports a causal relationship between the food or the property of the food and the health effect. This will determine whether the evidence dossier is compliant or non-compliant with the requirements of Standard 1.2.7 "*Nutrition, health and related claims*".

Some elements of Schedule 6 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. A number of different factors 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 based on balancing the impact and significance of different factors. In some cases, there may be differences between NZFS's judgements of the evidence and the systematic review author's judgements. To make the evaluation process as

transparent as possible, the additional guidance provided in this report will help ensure that the rationale underpinning the judgements by NZFS on the evidence provided, are consistent, reasoned and clear.

4 Set the scene for the systematic review

4.1 Develop the food-health relationship

To start a systematic review, the research question must be developed. The research question will guide the search strategy to find relevant studies to include in the systematic review. The review question must be designed to demonstrate the connection between the food or property of the food and the health effect (i.e., the food-health relationship). The food-health relationship underpins the self-substantiated general level health claim which must be phrased with equivalent meaning to the wording of the food-health relationship. It is important that the proposed direction of the relationship between the food or the property of the food is mentioned in this part of the systematic review: e.g., Phytosterols, phytostanols and their esters ***reduce*** dietary and biliary cholesterol absorption.

4.2 Define the food or the property of the food

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 several factors, such as whether the health claim is about a specific food product, a nutrient in that food, or another property of the food. A food business may need 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. The research scoping exercise will also determine which relationship needs to be assessed and what the comparator food/property will be. Note that in some cases, studies will not be able (or available) to provide evidence for the exact property that causes the health effect.

The food or the property of the food must be defined and characterised well. Certain properties of foods, such as an "antioxidant", will need to be characterised further before forming part of a food-health relationship. For example, vitamin C is an antioxidant which is a measurable food component.

Example 1: Whole plant food

Prunes

- Prunes could be defined as dried plums of "prune" cultivars (*Prunus domestica* L.).
- Nutrients relevant to the health claim will need to be quantified, which involves information on the method of analysis. For example, the methods for analysis of total dietary fibre is outlined in Standard 1.2.8 'Nutrient Information Requirements' of the Food Standards Code.
- State whether the food will be consumed raw or processed further, (and if it is processed, detail the method of processing).
- Where more than one part of a food could be eaten, the part that is the subject of the food-health relationship should be specified (e.g., skin, flesh, etc.)

Example 2: Property of a food

Probiotic

- To characterise a probiotic, the genus, species, and strain of a probiotic must be specified.
- The source of the probiotic should be described (e.g., product of fermentation of a specific food)
- Manufacturing process
- Genetic stability during production and storage
- Safety
- Dose or quantity present in the food (and how this is measured)
- Stability
    - Is the probiotic stable over time, or against heat or acid?
    - Is the probiotic stable within the matrix of the food?

Example 3: Property of a food

Amino acid (added to a supplemented sports food)

- The amino acid should be identified and described
    - Physical and chemical properties, composition
    - Naturally occurring sources
    - What are the methods to manufacture or isolate the amino acid?
- The food matrix should be described[1]
    - Consider how the added amino acid can be measured in the food
- Stability
    - Is the amino acid stable over time?
    - Is the amino acid denatured by heat or acid?
    - Is there variability from batch to batch?
    - What are the standard analytical methods for measurement?

4.3 Define the health effect

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.

In defining an effect on health, there is 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. Note this series of guidance documents are available here in the 'scientific guidance' section. It will also be helpful to consult someone with expertise relevant to 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 needed 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.

The *following is the 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.

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.

Only general level health claims are eligible for self-substantiation. 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.

Examples of health effects and health outcome measures

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 outcome 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) as the outcome measures.

A **functional process or outcome** could refer to an effect on skin barrier function (i.e. reducing the risk of skin dehydration). This outcome 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 as outcome measures.

Physical performance could refer to an effect on endurance capacity (exercising to fatigue). This outcome could be assessed 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 outcome 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 outcome could be assessed by measuring the occurrence of the common cold in a standardised way.

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 a non-serious illness such as 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.

The health effect claimed may differ from the outcome measures, or methods used to assess the health effect. For example, the health effect 'supports normal immune function', could be assessed by measuring the prevalence, duration, and severity of upper respiratory tract infection or acute gastrointestinal infection. *Reducing the severity of upper respiratory tract/gastrointestinal infection* would be considered a high-level health claim, but this outcome measure may still be considered appropriate to justify a general level health claim for maintaining a healthy immune system.

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 undertaking the search.

Example of information needed on property of the food and health effect.

Food health relationship: *"Whey protein has an effect on growth or maintenance of muscle mass"*

Property of the food = whey protein

- Which animal product is the whey protein sourced from?
- What dose is proposed to elicit the health effect?
- Which processing methods are used to separate the whey protein?
- What types of food matrices is whey protein likely to be used in, and are other components of those foods likely to interact with the action of whey protein?*

Health effect = growth and maintenance of muscle mass

- Which outcome measures are used to assess muscle mass growth / maintenance?
    - Examples include DEXA scanning, anthropometry.
    - Note that methods should be widely accepted or validated.
    - Provide details of the validated method, such as the make and model of the technology applied (eg, DEXA scanning machine, callipers).

* Note that to answer this question, it will be important to consider the types of studies that are used to provide the evidence to substantiate the food-health relationship. For example, relevant studies may include those which use whey protein in the final food product (e.g., a ready-to-drink protein drink), and studies which have used whey protein isolate as a supplement alone (provided appropriate comparators are used). If the evidence is primary or exclusively comprised of studies that test the effect of whey protein as an ingredient on its own, additional evidence will be required to show that a similar health effect exists for whey protein when it is consumed within a food matrix such as a beverage. This would be a demonstration of the relative bioequivalence of the property of a food when consumed in the food matrix. See the FSANZ guidance document for more information.

4.4 Develop the review question

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', so please also refer to this.

4.4.1 Determining PICOTS

Once the food or the property of food and the health effect have been established, the next step is to formulate the review question so that the search strategy for the systematic review can be developed. The document by FSANZ "Guidance on establishing food-health relationships for general level health claims" mentions the use of Participants, Intervention, Control, Outcome (Time and 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 PICOTS).

Example of developing a review question with PICOTS.

Food health relationship: *"Phytosterols, phytostanols and their esters reduce dietary and biliary cholesterol absorption"*

- **P**opulation/participants – generally healthy adults
  - Set an upper limit for blood pressure and cholesterol
  - Determine which medications qualify as exclusion criteria (i.e., cholesterol-lowering medication)
- **I**ntervention/**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).
- **C**omparison/control – placebo control, i.e., the food without the plant sterols/stanols.
- **O**utcome – dietary and biliary cholesterol absorption, calculated by measuring faecal cholesterol excretion
- **T**ime – ≥3 weeks duration
- **S**tudy design – randomised controlled trials*

* 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.

The participants in the studies should reflect the target population for the health claim. For example, studies where participants are children will not be eligible to support a health claim where the target population is adults (or vice versa).

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 at the start, rather than being added later. For example, consider whether participants will include athletes as well as non-athletes, and decide whether you will investigate whether fitness status influences the food-health relationship.

4.4.2 Appropriate study designs for supporting a food health relationship

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 minimise bias in the association of the food or property of the food with the health effect and show a causal relationship (Hariton and Locascio, 2018). Evidence from other types of studies (such as non-randomised controlled trials, cohort studies and case-control studies) can also be considered when judging whether a causal relationship exists, but they cannot be used alone as the only source of evidence to support a causal relationship. A greater weighting will be given to high-quality randomised controlled trials.

In the field of nutrition, cohort studies are commonly used to assess the effects of diet on the risk of developing certain types of non-communicable (and often serious) diseases. Sometimes secondary health effects are measured at the same time as dietary intake, and while these health effects may be appropriate for a general level health claims, this type of evidence cannot be used to support a food-health relationship. When information on health status is measured at the same time as diet is assessed, this is called a cross-sectional analysis because associations are identified at one time point only. A

food-health relationship must be supported by evidence of a *causal* association, which can only be demonstrated in a randomised controlled trial.

Case-control studies involve recruitment of cases (participants with a disease or condition), and controls (participants without that specific disease, but similar in every other way). Usually, a case-control study design is chosen because the disease or condition in question is relatively rare, and an observational trial would require an inordinately large number of participants in order to observe the outcome. Note that many case-control studies assess risk factors for *serious* diseases, and serious diseases cannot be referred to in a food-health relationship for a general level health claim applicable to the general population (because this would be a high level health claim as defined in section 4.3 of this report).

A common limitation of cohort and case-control studies is the imprecision of dietary assessment data. The method of dietary assessment is an important consideration in whether or not a study provides evidence of a food-health relationship. In large cohort studies dietary intake is often measured using a Food Frequency Questionnaire (FFQ), which assesses the intake of certain foods, usually at a grouped level (i.e. margarine) as opposed to individual foods (i.e., margarines containing phytosterols). Even if detailed dietary recalls have been used to measure food intake, these may have been collected at a time before the property of the food being assessed was developed.

5 Set the scope of the systematic review

5.1 Define the inclusion/exclusion criteria

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 these criteria should be 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 Thomas, 2022). In a systematic review, it is essential 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 will 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.

Below is an example of exclusion criteria that were used for a systematic review of phytosterols and cholesterol absorption. Note that most of the inclusion criteria relate to aspects of the study question that were outlined using PICOTS, as per the example listed in section 4.4.1.

Example of inclusion and exclusion criteria based on PICOTS

PICOTS	Inclusion criteria	Exclusion criteria
Population/participants	Generally healthy adults (specify age range and healthy blood pressure and cholesterol range)	Infants and children Adults outside of age or healthy blood pressure or cholesterol ranges, or who have a diagnosed illness that affects lipid metabolism (eg coronary heart disease, liver disease). People on pre-specified medications (eg, blood pressure lowering and cholesterol lowering medications). Animals
Intervention/exposure	Phytosterols, phytostanols and their esters (plant sterols, stanols and their esters) (e.g., sitosterol and campesterol - the most common form found in foods).	Studies where phytosterols, phytostanols and their esters (plant sterols, stanols and their esters) alone are not one of the interventions.
Comparison/control	Placebo control (i.e., the food without the plant sterols/stanols.)	No placebo/control group.
Outcome	Dietary and biliary cholesterol absorption, calculated by measuring faecal cholesterol excretion.	Did not measure cholesterol absorption as outcome.
Time/duration	≥3 weeks duration.	Study duration is not long enough to see a sustained effect (<3 weeks of receiving intervention).
Study design	Randomised controlled trials.	Not original research (i.e., reviews, editorials, letters to the editor).

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

When generating the search strategy for a systematic review, it is important to find a 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 Thomas 2022). This is why it is vitally important to identify detailed inclusion and exclusion criteria when reviewing a large number of identified studies. Librarians at Medical School/Health Science libraries have expertise in helping to choose relevant search terms, and it is recommended that you seek their help in developing your search strategy once the review question has been finalised. This section will provide some guidance on creating a search strategy, but it is recommended that you consult resources such as this article by Bramer *et al*, (2018).

There are two main components included in the search strategy: terms related to the food or property of the food and terms related to the health effect of interest (Higgins and Thomas 2022). Following is some guidance for conducting a search using PubMed, which searches MEDLINE – a freely available 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 several other subject headings listed below the search term in the hierarchical structure. If all 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 select 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 both free text and MeSH terms are incorporated into the search strategy. Using the advanced 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 recovered, 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. Below is an example of the way a search strategy could 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.

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). This is also useful to cover multiple accepted spellings of words (e.g., f*etal will pick up fetal and foetal).

5.3 Reporting the search strategy

The search strategy description should provide enough detail to allow NZFS to independently repeat the search. This information should include:

- The specific databases that were searched.
- The exact search terms and MeSH terms used, and how they were combined (AND / OR)
- The number of records obtained for each search.
- Any other relevant information, such as whether a date restriction was used (the reasons for applying such a restriction would need to be justified).

It may be useful to include screenshots of the search and the search outcomes.

6 Undertaking the literature search

6.1 Databases

Relevant databases to search are listed in the FSANZ guidance document. It is good scientific practice to search at least two of these databases. This will result in duplicate records, which can be excluded in the screening stage, but it is essential to search two or more databases because each database may identify different papers. Some databases (e.g., Scopus and Embase) also search conference proceedings. Databases may differ from each other in their format (e.g., MeSH terms may not be the same or present across all databases), so there may be minor differences in each search strategy. Any differences in search strategy between databases should be recorded.

It is useful to search the Cochrane Central Register of Controlled Trials (CENTRAL) (and/or a similar controlled trials registry), to determine whether there are any relevant trials not yet published. Often, completed trials may not be published immediately. Some trials may not be published at all because the results are not statistically significant or considered 'negative' (Song *et al,* 2000), and failing to include null results will bias the results of a systematic review. Searching a clinical trials registry will also identify whether any relevant studies are in progress, and this will help to indicate the appropriate time to update the systematic review.

The strategy used for searching clinical trial databases is quite different from the method used to search other databases, and so the search strategies will need to be modified accordingly. Searching for relevant studies in the clinical trials database should not be an onerous task since there are far fewer studies included here than in electronic databases of publications. If a registered trial has already been published, there should be a reference to the paper in the clinical trial database.

As well as searching databases and a clinical trials registry, it is recommended that the reference lists from included and relevant articles are hand searched to find any 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, Rayan, ProCite or EPPI-Reviewer. A database manager will help to keep 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 specific purpose of conducting systematic reviews and have additional features, such as being able to assign which criteria apply to the studies that are excluded. There is also the option to conduct a meta-analysis using the software within some of the databases (if it is appropriate to do so).

6.3 Identify relevant studies

Screening of the search results usually begins with a title screen, where studies that are clearly not eligible can be removed (i.e., animal studies, review papers, letters to the editor). This is followed by screening the abstracts of the papers that remain, and finally screening of the full text articles. It is recommended to list the number of initial search results, and the number of results after screening of title, abstract and full text. This information is usually presented in a flow diagram such as the example given in section 6.5 of this report.

It is helpful to record a summary of the reason(s) for exclusion of papers at each step of the screening process. At the title and abstract stage, a summary of reasons can be provided (e.g., *'not in humans; not randomised trials; irrelevant outcome'*); although if hundreds of titles and abstracts are screened, it may not be feasible to record and provide reasons for exclusion. However, at the full text screening

stage, it is best practice to list the excluded papers and the reason(s) for exclusion. This could be provided as an appendix. This list of excluded papers is particularly useful for NZFS when evaluating the systematic review, as other papers may be identified that should have been included. Without the list of excluded papers NZFS may not find it possible to determine whether these papers were identified and excluded by the original authors for a specific reason, or whether they were not identified in the original search.

Often, more than one of the exclusion criteria will apply to one paper, so a hierarchy could be applied when assigning an exclusion criterion to an article. For example, if a study is excluded due to meeting exclusion criteria 1, 3, and 5, then exclusion criterion (1) is assigned to that citation; or if exclusion criteria 3 and 4 apply, then exclusion criterion (3) is assigned to that citation. The authors of the systematic review can decide what hierarchy to use, if they choose to use one.

During screening, studies may be identified that are relevant for background information or to support a discussion. Sometimes a study may not meet the eligibility criteria for inclusion in the list of final studies as direct evidence for the causal food-health relationship, but it may provide important context for the discussion of the findings. For example, while an *in vitro* study about the mechanism of action of the property of the food on markers of health will not be eligible for inclusion, it may be useful to include it in the required discussion of biological plausibility (S6—2(f)(iii)). 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 systematic review.

6.3.1 Use of unpublished data

Even when the methods used to identify all published evidence as described above are appropriate, there may also be a bias in which studies get published, so it is important to assess whether the systematic compilation of literature yields the correct overall view of the evidence for a relationship. To help avoid this publication bias, you are encouraged to also search for unpublished data. Checking clinical trial databases can help identify completed studies that have not been published. Companies often also have unpublished evidence from in-house or pilot studies that should be considered for inclusion and to provide a more complete picture of the totality of the evidence.

6.4 Duplicate search recommended

As a quality assurance routine, it is recommended that at least two people complete the search strategy and screen the articles independently. The final lists of relevant studies identified may then be compared, and any differences or disagreements about eligible studies can usually be resolved by discussion or the input of a third person. This process will help to reduce the potential for bias in the systematic review methodology. Again, some software programmes can help manage the screening of records by all authors and automatically highlight where inconsistencies between different authors exist.

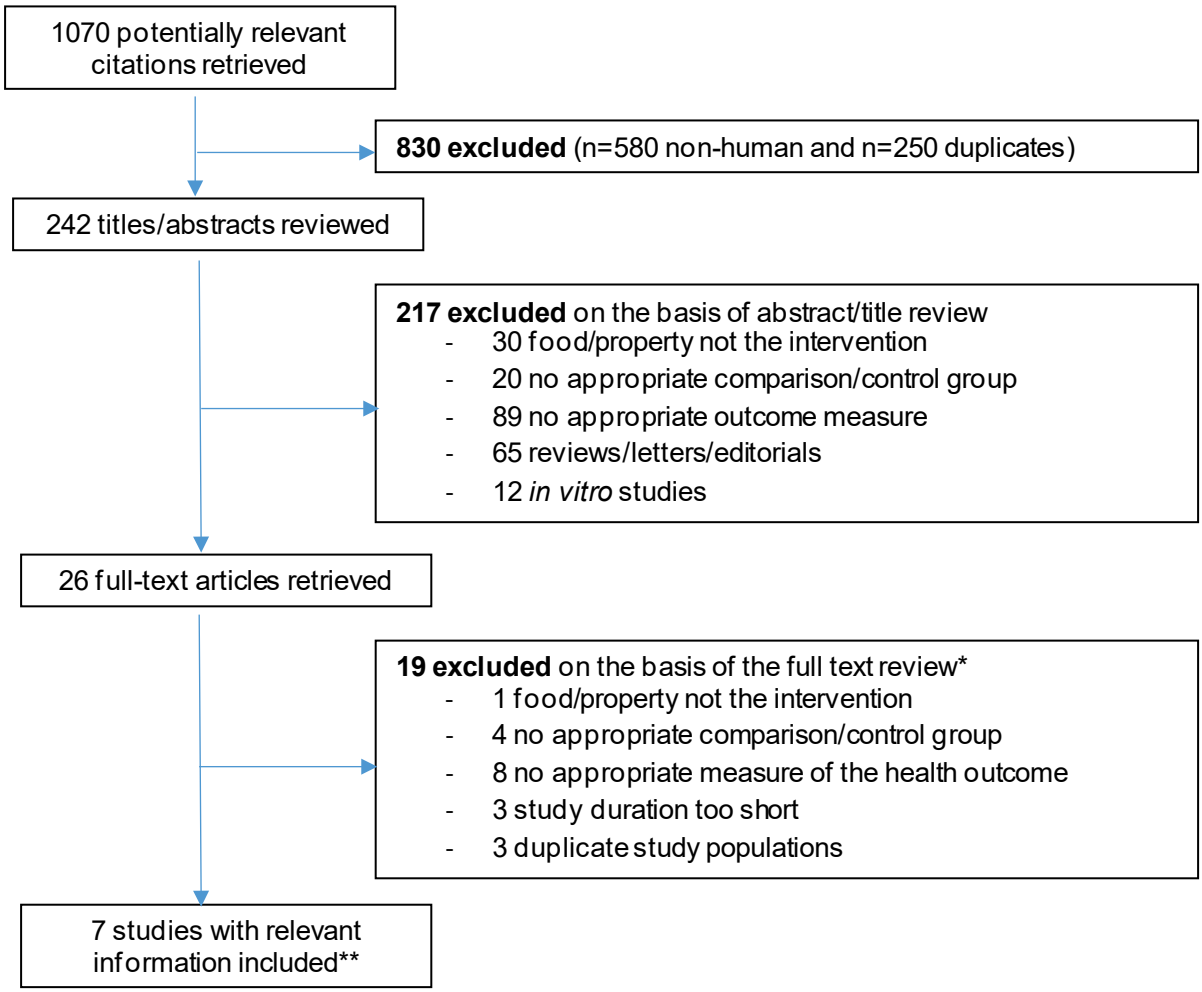
6.5 Finalise the list of studies included in the systematic review

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.

To summarise the processes described in **Section 6** of this report, a flow diagram of studies included in the evaluation of the evidence is a useful addition to the report (see **Figure 2** for an example). Information specific to the health outcomes and the 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.



*it is useful to tabulate these excluded references in an appendix with specific reasons for exclusion as this transparency allows reviewers to understand why studies they might identify as relevant were not considered relevant by the systematic review authors.

** The seven studies identified as eligible for inclusion must be listed, with full references, to meet the requirements of Schedule 6—2(c). All of these studies must be in humans.

7 Evaluate the evidence

7.1 Construct summary tables and extract data from the studies

After finalising the list of studies to be included in the systematic review, the key information from each of the studies must be tabulated.

The required elements of a systematic review are outlined in Schedule 6—2 (d):

> A table with key information from each included study. This must include information on:
>
> (i) the study reference
> (ii) the study design
> (iii) the objectives
> (iv) the sample size in the study groups and loss to follow-up or non-response
> (v) the participant characteristics
> (vi) the method used to measure the food or property of food including amount consumed
> (vii) confounders measured
> (viii) the method used to measure the health effect
> (ix) the study results, including effect size and statistical significance
> (x) any adverse effects.

Note that the above list is the minimum essential information, and it is suggested that reviewers add other items to the table if relevant. Some of the table's columns 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 or outcome. Participants' background diet and other lifestyle variables must be considered when assessing study quality, so it might be useful to document this information in the table alongside the other key information. An example of a study that has been tabulated according to these requirements in the guidance document is provided in the FSANZ guidance document (Appendix 2, page 26). Ultimately, the table with the key information from each study should be detailed enough that it can be used as the frame of reference for understanding the discussion section of the systematic review.

Item (vi) from the above list specifies that the method used to measure the property of the food, as well as the amount consumed, is listed. Many studies don't actually re-measure to confirm the doses used, but they should explain how the dose is prepared and delivered, who prepares the doses and how they determine how much of the dose is consumed. This should suffice as long as the method to measure the property of the food in each study is covered in S6—2(a).

For requirement (vii) "confounders measured", see section 7.2.4 of this report for some guidance on what to look for and include.

Where possible, make sure that the *numerical* results from the tables/figures in the studies are reported in the section on study results. It is not sufficient to describe results in general terms, such as by stating 'the difference between groups was statistically significant'. For example, results should ideally be reported as the mean difference between groups, with the 95% confidence interval and p-value. Information such as units of measurement should be included. If the authors of the original research study did not report a mean difference and 95% confidence interval (or an effect size), state in the table that this information is not reported. Report the relevant outcome measure results with as much numerical detail as the study authors provided. Note that it may be possible to obtain this information by contacting the original study authors. In addition to reporting study results in the study summary table, it might also be helpful to compile a summary table for study results only, which may be useful when considering whether there is a consistent association between the food and the health effect. Another option is to add the study quality rating (eg, high, medium or low) to this table once you have completed this.

Schedule 6 requires that any adverse effects caused by the food or property of the food must be documented. For example, if examining the effects of a food or property of a food on bowel function,

adverse effects would include both constipation-like symptoms (such as infrequent bowel movements and stools that are hard and difficult to pass), and diarrhoea-like symptoms such as very frequent bowel movements and stools that are watery and liquid-like. Each of these are 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. Note that adverse effects may not necessarily be related to the health effect – for example, symptoms of gastrointestinal discomfort may result from an intervention that is not expected to have an effect on the gastrointestinal system but should still be noted.

7.2 Assess methodological quality and applicability of each study

To obtain reliable conclusions from the systematic review and to comply with Standard 1.2.7, it is essential that results are interpreted 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:

(i) a clearly stated hypothesis
(ii) minimisation of bias
(iii) adequate control for confounding
(iv) the study participants' background diets and other relevant lifestyle factors
(v) study duration and follow-up adequate to demonstrate the health effect
(vi) the statistical power to test the hypothesis.

The list of study quality elements in Schedule 6—2(e) are the *minimum* elements that must be included in the quality evaluation for each study. The components considered in each quality evaluation depend on the study design and topic – for example, quality considerations for a study that undertakes dietary assessment must include a determination of whether dietary assessment was done using a validated process.

Some components of quality assessment will be more important than others when 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 particularly important when assessing the quality of cohort studies, because often there are several dietary factors assessed. However, if only randomised intervention studies are included in the systematic review, this component (though still required) may be less important to the overall quality rating compared with other quality indicators such as adequate statistical power to test the hypothesis or adequate control for bias. NZFS suggests that the assessment of all quality components should be described for each study included in the systematic review. This allows NZFS to understand the rationale behind the overall quality rating for each study included in the systematic review. It is recommended that at least two reviewers carry out the quality ratings separately, and have any differences resolved by discussion and consensus or by a third reviewer.

The following sections (7.2.1 – 7.2.7) will provide guidance for assessing each of the study elements listed in Schedule 6—2(e). Since Schedule 6—2(e) specifies that the subsections (i) to (vi) are the *minimum* required elements of study quality assessment, there may be other quality measures that are relevant to discuss. It is not feasible to identify every possible quality issue, but some of the more common additional quality issues will be covered in Sections 7.2.8 to 7.2.10.

7.2.1 Assess whether there is a clearly stated hypothesis

Required by Schedule 6—2 (e)(i)

A clearly stated hypothesis is important to determine whether the aims and objectives of the study are closely aligned with the food-health relationship that is the subject of the systematic review. This is particularly relevant for cohort studies that have multiple dietary variables that can be related to many different health outcomes (Smith and Ebrahim 2002). It is also relevant to intervention studies that have outcomes relevant to the food-health relationship which might not be a primary or secondary objective mentioned in the original study protocol. Ultimately, the original study hypothesis should be relevant to the food-health relationship because findings that are only present in *post-hoc* analyses (such as within a subgroup of the study population) are not considered to demonstrate a causal relationship as they may simply be due to coincidence (Curran-Everett and Milgrom 2013).

If there is no explicit hypothesis stated (that is, the predicted direction of the food-health relationship) in the publication, this could be found in a study protocol if one is available in a clinical trials registry. Note that sometimes the primary outcome reported in the trial register may differ to that reported in the paper. In this case it may be best to access the full study protocol to judge which information is correct. Otherwise, aspects of the hypothesis can be identified from key information in the study objectives or aims.

Finally, it is important to assess whether the study reported on all outcomes that were proposed to be measured. If there is incomplete or selective reporting of certain outcomes, this would bias the results of the study and would be a quality concern.

7.2.2 Assessing bias in intervention studies

Required by Schedule 6—2(e)(ii)

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 come from many sources:

Selection bias may occur in the way that participants are allocated to groups. If the randomisation procedure is flawed or if there is foreknowledge of the allocation sequence, this 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 by the research personnel.

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).

Note that bias is not the same as imprecision, methodological quality (e.g., is the study appropriately powered?), or inadequate reporting of results, although each of these do also influence the overall study quality.

7.2.2.1 Selection, Performance and Detection Bias

The key to avoiding selection, performance and detection bias is appropriate double-blinded randomisation of the allocation of participants to groups.

Reducing selection bias

Selection bias can be avoided by appropriately randomising participants to groups. If a study compares an intervention (food or property of the food) with a control or a comparator group, the participants in these groups should be as alike as possible in all aspects that matter for the health effect, before the study begins. Correct randomisation will minimise the risk of the groups differing in their key characteristics, such as mean age, sex, body mass index, or other relevant measures. Unmeasured confounding variables will also be assumed to be similar between groups, thereby minimising selection bias, though this is also dependent on an appropriate sample size (Haynes *et al,* 2012). Ultimately, if randomisation is done correctly, any differences between groups will be due only to random chance; and if the sample size is adequate, any differences between groups will be small.

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. Minimising selection bias helps to ensure this. Both groups of participants come from the same source and have been assigned to each group by random, preferably concealed, 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.

The following points provide examples of ways to identify and thus minimise or eliminate selection bias:

- The type of random sequence generation chosen will depend on the nature of the study, but could include simple or unrestricted randomisation, stratified randomisation, or minimisation (Cochrane Handbook 8.3.1). Most importantly, randomisation must not be predictable – so it would not be appropriate to randomise participants based on date of birth, day of visit, or to use alternate randomisation (i.e., where every second participant is allocated to the intervention group). Allowing participants or researchers to choose the group allocation would also not be random.
- *Allocation concealment* means that when a person is enrolled in the study, it is not possible to predict which group they will be allocated to. This way, the random sequence (if there is one) will be strictly implemented. For example, allocation is not concealed if the random sequence is known to researchers in advance, if the sequences are predictable (i.e., not random), or if sequences are held in unsealed or transparent envelopes. Poor allocation concealment could result in selective assignment of participants to groups (Pildal *et al,* 2005). When concealment of allocation is absent, intervention effects can be exaggerated, especially when the outcome is subjective (i.e., self-reported ratings of mood or pain) (Wood *et al,* 2008).
- The method of randomisation should be described, including details such as 'random allocation sequence was concealed'. Lack of detail on randomisation methodology is considered to increase risk of bias (Altman *et al*, 2001). If the method of randomisation is not described, it may be useful to look at the study protocol (in a clinical trials registry) to see whether there is more detailed information reported there.

Reducing performance bias

Performance bias is introduced when the researchers treat the intervention and control groups differently, beyond what is stipulated in the protocol. For example, in a study where exercise performance is an outcome, participants in the intervention group could be given more verbal encouragement from research staff compared with participants in the control group. Moreover, if participants are aware of their group allocation (such is often the case with nutrition intervention trials), they may inadvertently change their behaviour. These situations can lead to changes in the study outcome that are not due to effects of the intervention itself. To prevent performance bias, the study

must be double-blinded. This means that neither the participant nor the research personnel are aware of who is receiving the intervention or the control for the duration of the study. Research personnel in this situation refer to both those who allocate participants to the groups, those who are involved with measuring the health effect, and those who conduct the data analysis. Unblinding occurs *after* the primary analysis is completed.

A study that claims to be blinded should provide details of how the blinding was done (Altman *et al*, 2001). If a paper does not include a description of the method of randomisation, or if their description indicates that the allocation to groups was not random, the study is a "controlled clinical trial". These studies will still provide important information to support a food-health relationship, but the results will have to be interpreted considering the bias introduced by possible differences in confounding variables between the groups.

Not all studies can incorporate blinding. In many nutrition studies, blinding is not feasible if the intervention is very obviously different from the control. For example, in a study that compares whole-grain foods with refined grains, the foods consumed will be visually and texturally different. The degree of impact that this bias exerts will also depend on the outcome measured; for example, a subjective outcome such as satiety or hunger will be more likely to be affected by performance bias than an objective outcome such as blood glucose concentration. Evidence from a combined analysis of a number of meta-analyses shows that the absence of blinding leads 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 are relevant 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 equivalent placebo. It will be important to carefully consider the degree of impact that type of bias may have introduced in these studies.

Reducing detection bias

Detection bias can be avoided when those who assess outcome are blinded to knowledge of the intervention received. Sometimes, the outcome assessors are the same as the research personnel who enrolled participants, or the outcome assessors may be the participants themselves (e.g., in a study relying on self-reporting of symptoms). In that case, detection bias overlaps with performance bias. In other studies, the outcome assessors may be independent (e.g., clinicians who are external to the study), so detection bias is assessed separately from performance bias.

7.2.2.2 Attrition Bias

Almost all randomised trials will have some missing values for outcomes. Incomplete outcome data is known as attrition bias. This can occur when available data are not included in the report, or when there are systematic differences between intervention and control groups in the number of participants who drop-out or withdraw from the study. When this happens, the analyses will deviate from the intended randomised comparison (Higgins *et al,* 2011; Pocock and Abdalla 1998). For example, if many more people in the control group withdraw from the study, compared with the intervention group, there will be inadequate comparison data to demonstrate the effect of the intervention. Attrition due to non-compliance can lead to the effects of an intervention being over- or under-estimated (Pocock and Abdalla 1998).

There is no simple rule to determine how much missing data is too much. The degree of bias caused by attrition will depend upon the proportion of participants with missing data on the health outcome, and the extent to which there is an imbalance in missing data between the groups. It is also important to consider how the missing data are treated in the analysis. In a *per protocol* analysis, participants who are non-compliant with their treatment at any stage throughout the study are excluded from the analysis. Bias can be introduced if, for example, a larger proportion of participants do not comply with the intervention compared with the control. In this situation, the reasons for non-compliance (i.e., adverse effects related to the intervention) may affect the results. For this reason, often an *intention-to-treat* analysis is done, where the participants' last known data are carried forward and used as final values, despite an early withdrawal from the trial or other non-compliance. In some situations, missing data can be imputed, but it is essential that imputation is appropriate for the analysis and done with input from a biostatistician. Methods for the imputation should also be provided.

A key element of a randomised study is a description of the flow of participants through the trial (Altman *et al*, 2001). Usually, this is done in the form of a participant flow diagram. The flow diagram or

description will report the number of participants who were randomly assigned, received the intended treatment, and who were included in the analysis for the primary outcome. Reasons for withdrawal or non-completion should be given. If this information is not provided, this introduces attrition bias.

7.2.2.3 Reporting Bias

Reporting bias describes the selective or incomplete reporting of outcomes in a study. 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). Reporting bias can be difficult to immediately identify, but there are several ways that this can occur:

- Outcomes are measured (or likely to be measured) but are not reported in the results.
- Outcomes are reported in the results but are not mentioned in the methods.
- The full amount of data for the main outcome is not reported.
- Subgroup analyses are reported alone, without the primary analysis.
- Adjusted analyses are reported alone, without an unadjusted analysis to compare with.
- Results are reported in insufficient detail[1]
    - e.g., results are described as 'significantly different' or 'not-significantly different' with no numerical results shown.
    - e.g., only p-values are reported, but no effect size or mean difference is reported.
- Statistically significant results are over-emphasised while non-significant results are simultaneously under-emphasised.

Ideally, a study will have a protocol published before the study commenced, where all methods, planned sub-group analyses and main outcomes are pre-specified. Some older studies will not have a published protocol, and in these cases, the methods and results should be examined to determine that the pre-specified and expected outcomes of interest are all reported in full. Sometimes, with large studies measuring multiple outcomes, results are published separately for groupings of outcomes. Reporting of results should include estimated effect size and precision (such as 95% confidence interval), based on participant comparisons (Altman *et al*, 2001).

7.2.3 Assessing bias in observational studies

Required by Schedule 6—2(e)(ii)

Due to their design, observational studies cannot be used to establish causality. However, observational studies may still be used to partially support a food-health relationship, especially where randomised controlled studies are not feasible or ethical, or where only very randomised controlled trials have been reported. Observational studies reveal associations between an exposure (i.e., consumption of certain foods or nutrients) and outcome(s). However, there remains the possibility that the association is not solely caused by the exposure, but rather a chance finding (i.e., if the study population is not representative) or due to bias (i.e., confounders exerting an independent effect on the outcome), or simply chance. Assessing the risk of bias in observational studies will help to establish the degree of confidence in the association between exposure and outcome (Hammer *et al,* 2009).

7.2.3.1 Selection Bias in observational studies

In a cohort study, it is important that exposed and non-exposed[2] participants are drawn from the same population. For example, it would not be appropriate to recruit only the 'exposed' population from one city and the non-exposed population from another city.

[1] Note – sufficiently detailed results can be entered into a meta-analysis. If this is not possible, (or only possible with imputation), there is risk of reporting bias.
[2] "Exposed" refers to the exposure of interest. I.e., the exposure of interest might be intake of a specific food or property of a food such as intake of coffee or dietary fibre.

Selection bias occurs when the study population is not randomly selected from the target population for which the finding will be generalised – for example when subjects volunteer because they have a high chance of exposure or risk of the outcome of interest (e.g., in a study on bowel cancer and dietary fibre, they have a high dietary fibre intake, or a high risk of developing bowel cancer).

It is also an important quality parameter (related to selection bias) to consider whether the participants recruited into the study are representative of the general population. For example, a group comprised only of doctors and nurses may be representative of the population of doctors and nurses, but not the general population.

7.2.3.2 Accuracy of dietary assessment

Dietary assessment methodology is of particular importance for an observational study. It is imperative that the dietary assessment method is validated – for example, a validated food frequency questionnaire (FFQ), or a 24-hour diet recall interview administered by a dietitian may both be appropriate, depending on what aspect of diet is being measured. Note that a validated FFQ must be validated specifically to the population that it is used on. Ideally, diet should be assessed on more than one occasion, since dietary patterns often change seasonally or over time. Usually, the outcome of interest will take many years to develop, so dietary assessment at multiple time points will more accurately link dietary patterns to the outcome. Measuring nutritional biomarkers in addition to assessing diet itself will also strengthen the dietary assessment. For example, a validated and repeated FFQ to assess iron intake would be strengthened by also measuring haemoglobin or serum ferritin levels, (even if this is done in a random subgroup of participates for practical reasons).

7.2.3.3 Outcome of interest confirmed as not present at recruitment

The outcome of interest (health effect) should be confirmed to be not present in the study subjects at the start of the study (i.e., the dietary pattern precedes the health effect).

7.2.3.4 Adequate follow-up time

Adequate follow-up time for the specific health effect should be allowed – often this will take a number of years. It may be necessary to justify the choice of follow-up time allowed.

7.2.3.5 Appropriate assessment of outcome

Assessment of the health effect (outcome) will ideally be confirmed by medical record linkage or by independent assessment by a (blinded) health professional; however, methods such as self-report are often used for practical reasons.

7.2.3.6 Attrition bias in observational studies

Some participants will inevitably be lost to follow-up in long studies, but the attrition rate should be reported and should be minimal. A large attrition rate could bias the results, since participants who are lost to follow-up may have different relevant characteristics than those who remain, (i.e., socio-economic status, dietary patterns, etc.).

7.2.3.7 Adequate control for confounders

Control for confounding factors is essential in a cohort study as it is a common source of bias. Often, results will be presented from an unadjusted model, a minimally adjusted model, and a fully adjusted model. There are some confounding factors that should always be adjusted for (including but not limited to age, socio-economic status, smoking status), and other relevant confounding factors will be dependent on the dietary exposure and the health effect. A confounding factor is chosen if it is likely to have an effect on the health outcome that is independent from the dietary factor being measured. For

example, a longitudinal study may assess the extent to which fruit and vegetable consumption is associated with cardiovascular disease. If a high consumption of fruits and vegetables is also associated with a more physical activity, physical activity will be a confounder that exerts an independent effect on the outcome (cardiovascular disease); and the effect of fruit and vegetable consumption on cardiovascular disease will be exaggerated.

It is always possible that there are unknown and unmeasured confounders that exert some influence on the outcome of interest, which will introduce bias to the findings. Further discussion on confounders is in Section 7.2.4.

7.2.4 Assessing whether there has been adequate control for confounding

Required by Schedule 6—2(e)(iii)

Assessing whether there has been adequate control for confounding is part of assessing the risk of bias in a study, though it is also an additional quality parameter required by S6—2(e)(iii).

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.

7.2.4.1 Control for confounding in randomised trials

In randomised trials that have achieved adequate randomisation (i.e., an appropriate method of random allocation to groups has been well described and implemented), this is considered to adequately control for confounding (Higgins *et al,* 2011). This is because any differences between groups can only be attributed to random chance, and any differences are likely to be small if the sample size is adequate. Most often, if randomisation is done correctly and the sample size is adequate, there will be no differences in participant characteristics between groups. Conversely, if the method of randomisation is not adequate or not described, the risk of bias due to confounding is higher.

If a trial has a small sample size, it is more likely that baseline characteristics may differ between groups, even when randomisation is done appropriately. This may cause problems for nutritional trials due to variations in background dietary intake by chance. Therefore, baseline characteristics for both groups should always be reported. Despite any random differences, primary analysis should not involve control for baseline characteristics, as a general rule. Instead, a *post-hoc* analysis that measures the effect of the potential confounding factor could be useful. The exceptions to this are: if a baseline characteristic is a known confounder, and the decision to adjust is made *a priori* and included in the trial protocol; or if there is likely to be a problem with the randomisation. However, if there is incorrect randomisation, adjusting for imbalanced baseline characteristics will not improve the study quality due to poor randomisation (Roberts and Torgerson, 1999).

7.2.4.2 Control for confounding in observational studies

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 diagnosed 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 those containing omega-3 fatty acids.

There should also be an assessment to check whether all important confounding variables have been included in the statistical model as covariates in an observational analysis. After adjusting for a confounding variable, it is important to check that all the effects of that variable have been removed. For example, if the continuous variable Body Mass Index (BMI) might confound the relationship between a food or property of a food and an outcome, it is essential to consider the method of adjustment for BMI in the statistical model. Adjusting for BMI as a categorical variable (two or more categories) might not capture all the necessary information required to achieve sufficient statistical control. This may result in inadequate control of confounding, which is 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 accordingly.

" Over-adjustment" for confounding variables should also be assessed, as over-adjusting may result in an underestimate of the effect of the food or the property of the food on health. If a variable mediates the relationship between the food and the health effect, it may not be appropriate to adjust for this. 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

Required by Schedule 6—2(e)(iv)

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 of the food) 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 should also be considered in the context of Australian and New Zealand populations, in a later section of 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

Required by Schedule 6—2(e)(v)

The exact duration of a study that is required to demonstrate the effect of a food or a property of a food on a health effect is dependent upon the proposed effect on health. For example, if the health effect relates to acute postprandial health effects, studies with duration of 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. However, health effects such as cognition and growth or bone density will likely require longer periods to identify changes (i.e., 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 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, and provide rationale for their decision.

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

Required by Schedule 6—2(e)(vi)

Published studies should report a sample size or power calculation (Altman *et al*, 2001). This involves identifying the primary outcome used to do the power calculation, the quantities used in the calculation, and the target sample size in each comparison group. The outcome that is used in the power calculation to determine the sample size is known as the *primary outcome*, but a study may also include several other secondary outcomes. If the outcome relevant to the food-health relationship is a secondary outcome (and therefore not reflected in the power calculation) there is a risk that the sample size may be too small to have confidence in the result. This may result in the quality of the study being reduced (for the purpose of the systematic review question).

Many randomised trials unfortunately do not report a power calculation, though it is worth looking up the study protocol for this information if it is available in a clinical trial registry. Note that a power calculation must always be done before the study is started.

If a study is underpowered, it is possible that the study will fail to reveal a true difference between treatment effects. Alternately, if an underpowered study reports a significant and large effect, it may still 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.

7.2.8 Recognising random measurement error

One of the main reasons to require a suitable control group within an intervention trial is to test what would happen over the duration of the study if participants received no intervention or received 'usual diet'. This is important in order to rule out the 'regression to the mean' phenomenon where measurements of multifactorial health outcomes with an element of chance can show variation 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 an extremely low immunological reaction based on accepted markers of immune function) who may be given an intervention to increase their immune response. When these participants have their immune system markers remeasured, they are likely to be closer to the mean of group, due to this common statistical phenomenon. Conversely, participants with an extremely marked immunological reaction at one measure are likely to show a decrease towards the mean upon re-measurement). This should not be interpreted as showing an effect of the intervention because even if participants are not treated the mean response will differ, owing to regression towards the mean. Therefore, having a suitable control group that have their markers of immune function 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 check 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 is not likely to 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).

Generally, it is important to recognise when regression to the mean is at play, and control for it as much as possible, as this can help avoid the misinterpretation of data and potentially seeing patterns that may not exist.

7.2.9 Using appropriate statistical analysis (Intervention studies)

An important quality parameter is whether statistical analysis of the data is appropriate for the study design, or whether a between-group statistical analysis of the health effect was conducted. The study design, method of data collection, and nature of the data being collected will all determine which statistical analysis method should be used. If the statistical analysis methodology is not reported in adequate detail, or if it is not appropriate for the data, it may be possible to contact the original study authors to obtain this information.

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: some specific study designs will require repeated measures Analysis of Variance (ANOVA)). A statistically significant *within-group* difference over time in the intervention group does not demonstrate a causal food-health relationship. Comparing a baseline measurement with a final measurement separately in each group is not appropriate for randomised controlled trials and will likely produce misleading results (Bland and Altman 2011). To properly test the effect of an intervention in a group of people, the effect of the intervention must be directly compared with the effect of the control.

As well as considering statistical significance in study results (usually demonstrated by $p<0.05$), the precision of the effect of the intervention should be reported. This is usually done by reporting the mean difference and 95% confidence intervals, which indicates 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 large sample size or that 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. There are other ways to communicate the magnitude of results, such as reporting effect size or standardised mean differences. It is recommended that someone with expertise in statistical analysis is also consulted to assess the appropriateness of the analysis in the quality assessment of each study.

7.2.10 Validity of the health outcome measures

Quality appraisal should also consider the reliability of the data collection tools used. This can apply to both the methods used to assess the exposure (e.g., dietary assessment techniques) and the methods used to measure the health outcome. It is important that the methods used to measure both the exposure

and 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 an exposure or outcome, the paper will usually state whether the questionnaire has been validated. Note that findings from a validated questionnaire are robust only when the chosen statistical methodology is appropriate and they are administered in full, as intended. It is useful to refer to the paper that demonstrated the validity of the questionnaire, since instructions on how to handle the data are usually included.

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 (note that 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. For this reason, NZFS do not recommend the use of any one specific tool. Note that the six quality aspects described in Schedule 6—2(e)(i - vi) are the minimum quality assessment requirements, and often there will be additional relevant quality issues to discuss, as described in section 7.2 of this report. All six quality aspects of the studies included in the systematic review must be assessed in order to meet the requirements of Schedule 6.

Some study quality appraisal tools use a scale or assign numeric 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 because 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, if any tool is used, it is recommended that at least two reviewers complete the chosen quality appraisal tool and compare their ratings.

Most tools assign a point of equal value to each quality component, regardless of the nature of the study (e.g., listing inclusion criteria and having an adequate randomisation method are given equal importance). While this is objective, it can also lead to inaccurate results if points are missing for the most important quality indicators. In addition, dividing studies into either "higher" or "lower" quality, based on the total score, tends to create an arbitrary dichotomy, when quality is more of a spectrum. Studies with a score at the lower end of "higher quality" can often be very different to studies with a score at the upper end of "higher quality". Therefore, while quality assessment tools are a very useful starting point to examine important quality indicators, **when NZFS evaluates systematic reviews, a wide range of quality indicators and limitations for each included study is assessed. These are discussed in terms of their relative importance to one another and the study type. NZFS then assigns a quality rating of low, medium or high to each study based on a detailed discussion of any study limitations. Therefore, using various quality assessment tools in preparation of your systematic review can certainly help, but more discussion and consideration of quality limitations and their relative importance to the level of certainty of the individual study results is encouraged.**

Observational or cohort studies are subject to different sources of bias than intervention studies. The tools that can be used to assess bias in these types of studies include (but are not limited to) the Newcastle Ottowa Assessment Scale (NOS), Effective Practice and Organisation of Care (EPOC) and the Effective Public Health Practice Project (EPHPP). The NOS assesses bias in non-randomised case-control and cohort studies, while the EPOC model grades the quality of health system interventions for systematic reviews in this area, and the EPHPP model assesses the quality of quantitative studies. This section is not a comprehensive description of quality appraisal of observational studies, but will outline several examples of bias specific to cohort studies.

7.3.1 Health Canada's Quality appraisal tool for intervention studies

While NZFS does not use this tool when evaluating systematic reviews, the quality appraisal tool for intervention studies from Health Canada was designed specifically for the assessment of the quality of studies included in a systematic review for the substantiation of health claims to meet Canadian regulatory requirements (Health Canada, 2009). There are fifteen quality 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 is given if it is not present (or not reported). A total score of eight or more means that the study is considered "higher quality", but if the score is less than eight, the study is considered "lower quality". Note that a "higher quality" study according to the Health Canada tool might not necessarily be considered a "high quality" study according to Schedule 6. The Health Canada quality appraisal tool covers two aspects of study quality (bias and confounding), but not all required aspects of quality that are outlined in Schedule 6 of Standard 1.2.7. Therefore, if the study is categorised as being "higher quality" this can apply to the bias and confounding aspects of study quality, but may not apply to the other aspects of study quality specified in Schedule 6 as a minimum (i.e., background diet and lifestyle factors, appropriate duration/follow-up, sufficient statistical power). These quality measures must be considered in addition to those in the Health Canada quality appraisal tool, preferably together with any other important quality considerations not assessed by the tool.

7.3.2 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 twelve 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, so the other essential quality appraisal measures must be added to the overall quality assessment.

7.3.3 Cochrane Collaboration tools 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, to indicate whether the results of the overall systematic review, are likely to be significantly affected by bias or not:

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

It is important to note that bias is only one of the quality indicators required by Schedule 6—2(e), so the Cochrane collaboration's tool for assessing risk of bias can be used to contribute to the study quality appraisal, but does not stand alone.

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 certainty of the evidence.

7.3.4 Summary on using quality appraisal tools

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 in 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 Schedule 6 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 7.2.1 and 7.2.5 to 7.2.7 for more information). The Cochrane Collaboration's tool assesses the risk of bias but other components of study quality covered in Section 7.2 of this report will still need to be assessed separately from this tool, and factored into the overall rating of study quality.

When NZFS evaluates systematic reviews, a wide range of quality indicators and limitations for each included study is assessed. These are discussed in terms of their relative importance to one another and the study type. NZFS then assigns a quality rating of low, medium or high to each study based on a detailed discussion of any study limitations.

Therefore, using various quality assessment tools in preparation of your systematic review can certainly help, but more discussion and consideration of quality limitations and their relative importance to the level of certainty of the individual study results is encouraged.

8 Assess methodological quality of studies as a group

After assessing the quality of each study, the higher quality studies should be assessed as a group to determine whether there is a consistent and causal association between the food and the property of the food.

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

An assessment of the results of the studies as a group by considering whether:

(i) there is a consistent association between the food or property of food and the health effect across all high quality studies
(ii) 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)
(iii) the proposed relationship between the food or property of food and the health effect is biologically plausible
(iv) 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. It can be difficult to evaluate the consistency of an association, especially for a health effect that is not measured using the same health outcome measures across all studies.

To be described as "consistent" (as required by Schedule 6—2(f)(a)), the effect of the food or property of food and health outcome must be demonstrated in more than one high quality study. 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). 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 study size, effect size, and the precision of the study results. Each of these factors will influence the level of confidence in the outcome, and therefore the likelihood of a new study having a different result.

8.1.1 Meta-analyses

It is **not a requirement of Schedule 6** to conduct a meta-analysis (statistical method that combines all of the data on the health outcome across the studies), but if it is possible and useful to do so, then this is a straightforward option to assess the consistency of the association across studies. Sometimes NZFS will conduct a meta-analysis to assist with the evaluation process, when it is appropriate do so. Note that a meta-analysis should only be done to combine data that are comparable (i.e., the same health outcome measure is used among similar types of participants, and from high quality studies). Combining data from poor quality studies is not recommended because this will not result in any increased confidence in the results.

It is recommended that a statistician is consulted before combining all the results into a meta-analysis as there are several assumptions underlying the different ways to combine the data, which can affect the overall results. Free access statistical software packages for conducting a meta-analysis are readily available (e.g., Revman by Cochrane).

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. Therefore, if the data are appropriate for meta-analysis, conducting a basic meta-analysis is preferable to NZFS and is also likely to be easier.

8.1.1.1 Measuring publication bias in meta-analyses

Although also not required by Schedule 6, when meta-analyses are undertaken, the most common and simplest way of detecting publication bias is to generate a funnel plot. This is a scatterplot of the estimate of effect from each study in the meta-analysis against a measure of its precision (e.g., standard error). If the resulting plot is a symmetrical funnel-shape, this indicates an absence of publication bias. If asymmetrical, this will be caused by smaller studies reporting negative results being missing – indicating a publication bias to be present (Peters *et al*, 2006).

8.2 Causal association between the food and the health effect

The guidance document by FSANZ (pages 14 and 15) provides useful information on 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. The key criteria include:

- Consistency of results
- Strength of association
- Dose response relationship, and
- Temporality.

Because consistency is already discussed in section 8.1, the remaining criteria are discussed below. It is important to note that these criteria for causation were developed for 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. A high quality randomised controlled trial (RCT) is the only study design that can unequivocally demonstrate a causal relationship between the intervention and the outcome. However, for some research questions, results from RCTs can be supported by results from observational studies.

8.2.1 Strength of the association

The greater the strength of an association (effect size, or magnitude of effect) 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 can be considered as a 'strong' association; however, if results show an effect on the health outcome in an order of 40-50% then this would probably be considered strong (Potischman and Weed 1999). Effect sizes may be used where appropriate to determine whether an effect is small, medium, or large (Sullivan and Feinn, 2012).

Often smaller published studies 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 will 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.

8.2.2 Dose-response relationship

A dose-response relationship exists when the magnitude of the effect is directly related to the amount of the food or property of the food consumed. A dose-response relationship may strengthen the likelihood that there is a causal association between the food/property of the food and the health effect, but absence of a dose-response relationship does not exclude the possibility of a causal association. It might not be possible to measure 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 into 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 7.2.8 'Recognising random measurement error' for more information).

8.2.3 Temporality

Causal relationships involve the concept of temporality, in that one action promotes a causal reaction over time. In determining whether there is a causal association between a food or a property of a food and a change in health outcome, the exposure (food) must be measured before the health outcome change (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. However, 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 whether altered intake of that food is a result of an effect on health (Hill 1965). For example, if a high intake of diet drinks is observed in subjects who are overweight, it does not mean diet drinks cause overweight. Perhaps rather it means that overweight people are more likely to consume diet drinks to help manage their weight. Therefore, evidence from studies lacking the aspect of temporality will not be able to substantiate a causal effect.

8.3 Biologically plausible association between the food and the health effect

Biological plausibility for a food-health relationship must be discussed as part of establishing a causal relationship. This is required by Schedule 6—2 (f)(iii). Studies that might contribute to 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 acute post-prandial effects) might provide some useful information on the biological mechanism(s) that underlie the association being reviewed.

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 this effect 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 animal or *in vitro* studies demonstrating biological plausibility will not strengthen the relationship or demonstrate causality in humans.

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

Schedule 6—2(f)(iv) requires that consideration must be given to whether "the amount of the food or property of the food to achieve the health effect can be consumed as part of a normal diet of the Australian and New Zealand populations".

While 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, this required discussion is not part of the systematic review – it addresses the practicality of the food health relationship in the context of the overall diet.

This requirement of Schedule 6 is not a typical requirement of systematic reviews and therefore often needs to be provided in separate documentation from the food business selling the final product, rather than by the author of the systematic review.

Based on the effective dose of the food or property of the food it is important to discuss the viability of that dose within the final product and within the diet of the Australian and New Zealand populations. In the FSANZ guidance document there is a detailed description of factors to consider 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. For instance, if the health claim relates to 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). However, this may still be considered possible and appropriate according to the official Dietary Guidelines of each country.

NZFS recommends this discussion should include:

- how the final food for sale making the proposed health claim fits within the New Zealand/Australian dietary guidelines,
- research on quantities of consumption of the food category in question by Australians and New Zealanders, and
- whether consumption of the new proposed food will displace any other foods and therefore influence nutritional quality of the overall diet.

Because this is the area most commonly missing or inadequately completed in evidence dossiers evaluated by NZFS, some examples of these types of considerations for whole foods and a property of a food are provided below. It is recommended that assistance from a nutritionist or dietitian may be helpful in completing this aspect of the regulatory requirements.

Example 1: Whole plant food

Prunes

- Prunes would be grouped with dried fruits in the dietary guidelines for both countries. The New Zealand the Eating and Activity Guidelines (Ministry of Health, 2020) encourage an increase in intake of fruits to at least two serves a day. However, as prunes are dried, they are classified as 'high in sugar' and recommended to be limited in the diet due to their effect on dental caries. Therefore, if the evidence for an effective dose requires eating them daily, or several times a day, this may go against dietary recommendations unless they are to be consumed with a meal, rather than as a snack on their own.
- National nutrition survey data or specific dietary assessment data from other New Zealand or Australian food intake studies may be available to indicate average consumption of prunes in the population, or specific target population. This figure could be compared with recommended intakes from dietary guidelines and with the effective dose for a health effect. This discussion makes it clear how effectively an effective dose could be incorporated into the diet. For example, if the average consumption of prunes in the target population is two prunes a week, and the effective dose is ten prunes a day, it is easy to see that this will involve a significant dietary change, that may not be supported by the dietary guidelines. However, if the reverse is true, and the usual consumption is close to the recommended effective dose, it is likely to require minimal dietary change for the population and can be easily incorporated within existing dietary guidelines into meals to help protect dental health.
- Finally, it is necessary to discuss how incorporating the effective dose of prunes may complement, replace or add to overall nutritional intake at a population level. If ten prunes a day are required, this will add significantly to total sugar, energy and fibre intake. This can be quantified and discussed. Additional sugar and energy intake is not consistent with dietary recommendations, while additional fibre intake is. Therefore, it is useful to discuss what high sugar and energy foods prunes could replace, to provide the additional fibre without providing extra sugar and energy.

Example 2: Property of a food

Probiotic

- Unlike the example of a whole food above, a discussion on how a probiotic ingredient within a food can be consumed as part of a normal diet of the Australian and New Zealand populations is not possible without understanding what the final food product (or at least food category) is.
- If the final food product is a fermented milk or yoghurt, consideration must be given as to whether the effective daily dose can be delivered in one or two (or more) serves.
- As per the above example, the place of dairy products within the dietary guidelines can be discussed.
- As per the above example, the current intake of dairy products, fermented milks or yoghurts in the whole population or the specific target group within the population (e.g., children aged 1-3 years) should be discussed in relation to the recommended effective dose, in serves per day. If current intakes of dairy products or key nutrients such as calcium are inadequate, this product could help address existing nutritional needs in the population. The reverse may also be the case.
- Finally, as per the above example, if a specific effective dose in serves per day were to be added to usual intakes, what effect would that have in terms of overall nutrient/energy intake and what foods could be displaced by incorporating the requires serves per day? Would it be proposed to replace existing intakes of milk/yoghurt, or other non-dairy foods such as biscuits or cakes?
- Clearly if the probiotic were not in a dairy product, but in a cereal bar, or a breakfast cereal, or a non-dairy beverage, the above points and discussion would be very different.
- If a systematic review for a property of a food is compiled by an ingredients manufacturer the above information for the final food is not usually provided. However, it is very important that this context is discussed as part of the dossier of evidence to support a food-health relationship, as it is a key requirement of Schedule 6, and of the overall conclusion of the systematic review, as outlined in section 9 of this report.

9 Conclusion of the systematic review

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

> A conclusion based on the results of the studies that includes:
>
> (i) 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
> (ii) where there is a causal relationship between the food or property of food and the health effect:
> (A) the amount of the food or property of food required to achieve the health effect
> (B) 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.

No new information is presented in the conclusion; rather, Schedule 6—2(g) lists the content that must be included in the conclusion of the systematic review to establish the food health relationship.

Based on the discussion of causality in the previous section, the conclusion must be informed by whether or not a causal relationship has been established, based on the totality of the evidence. 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.

The requirement for a conclusion to include information on "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", can draw from the detailed discussion covered in section 8.4 of this report.

10 Updating an existing systematic review

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

An existing systematic review may be used if it is updated to include:

(i) the required elements of Schedule 6—2 (a) to (f) for any relevant scientific data not included in the existing systematic review; and
(ii) the required element of Schedule 6—2 (g) incorporating the new relevant scientific data with the conclusions of the existing systematic review.

If there is an existing systematic review that applies to the notified food-health relationship, this can be updated. This means that the original authors' search is done again to include the years between publication of the systematic review and the present day. Any new, relevant studies are then included.

If you choose to update an existing systematic review, the research question must precisely match the food-health relationship. For example, if the food-health relationship is "*Whey protein has an effect on growth or maintenance of muscle mass"*, an existing systematic review that investigates the effect of all protein consumption, rather than whey protein specifically, would not be suitable. A systematic review that limits participants to a specific subgroup, such as elderly people, will also not be suitable for self-substantiating this food-health relationship, since studies among other population groups would be omitted. Please review Sections 4 and 5 of this report in deciding whether an existing systematic review is suitable to use in substantiating your specific food-health relationship.

The remainder of this report will also be relevant when incorporating new studies and updating the existing systematic review. This is particularly important because most, if not all, existing systematic reviews will not include every requirement of Schedule 6. Significant new work will likely be required, especially with respect to quality assessment (section 7 of this report) and a discussion of the overall effect on the diet of the population (section 8.4 of this report).

11 Developing a health claim from a self-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 Schedule 4 of Standard 1.2.7 is prescriptive for the wording of the health claim, so any statement or information may be modified, provided it does not alter the meaning of the pre-approved relationship. The same is true for self-substantiated food health relationships and subsequent health claims.

Requirement of Standard 1.2.7—20(2):

> If a health claim is a general level health claim based on a relationship that has been notified under paragraph 1.2.7—18(3)(b), the health claim must:
>
> (a) state the food or the *property of food and the specific health effect; and
>
> (b) include together with the health claim a statement about the relevant population group, if any, that is a reasonable conclusion of the systematic review mentioned in paragraph 1.2.7—18(3)(b); and
>
> (c) include, together with the health claim, the information referred to in subsection (3).

It is therefore important that the wording of the health claim for an established food-health relationship includes 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 self-substantiated food-health relationship. It is important that the wording of the health claim does not convey an effect on health that is broader than the food-health relationship and does not take the health effect to a level beyond what the scientific evidence has demonstrated.

Some examples are discussed here:

Example 1: healthy immune system claim

Self-substantiated food-health relationship established for: "[x] is necessary for normal immune system function".

Equivalent wording within claim: "contains [x] for a normal/healthy immune system". ✓

Non-equivalent wording within claim (not permitted): "contains [x] to boost your immune system". ×

Rationale: The non-equivalent claim amplifies 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.

Example 2: digestive health claim

Self-substantiated food-health relationship established for: "[x] is necessary for normal laxation".

Equivalent claim: "contains x to help keep you regular / for regular bowel movements".✓
Non-equivalent claim (not permitted): "contains [x] for good digestive health". ×

Rationale: The non-equivalent claim is too broad. Regular laxation is only one aspect of good digestive health.

11.1 Ensuring all other required elements of a health claim are included

Other key elements that must be included in a claim (in addition to the food-health relationship) are outlined below.

Relevant requirements of Standard 1.2.7—20(3)-(6):

For paragraph 1.2.7—20 (2)(c), the information is:

(a) a dietary context statement that complies with subsection (6); and

(b) a statement of the form of the food to which the health claim relates.

(4) Despite paragraph (3)(a), a dietary context statement need not be included on a label on a food for sale that is contained in a small package.

(5) Despite paragraph (3)(b), if the form of the food to which the claim relates is the food as sold, the form of the food to which the claim relates need not be stated.

(6) A dietary context statement must:

(a) state that the health effect must be considered in the context of a healthy diet involving the consumption of a variety of foods; and

(b) be appropriate to the type of food or the property of food that is the subject of the claim and the health effect claimed; and

(c) (ii) if the health claim is a general level health claim based on a relationship that has been notified under paragraph 1.2.7—18(3)(b)—include words to the effect of a relevant dietary context statement that is a reasonable conclusion of the systematic review.

Therefore, 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. If the evidence in the systematic review is all on a specific population group (such as children or pregnant women), the wording of the health claim must also specify that the effect relates only to that population group.

NZFS recommends the dossier required by Schedule 6 contains information about the likely full wording of a health claim on the final food for sale. This will ensure the evaluation report written by NZFS provides optimised feedback to the food business making the self-substantiated claim, in relation to compliance with all regulatory aspects involved in making a self-substantiated general level health claim.

12 References

Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T; CONSORT GROUP (Consolidated Standards of Reporting Trials). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med., 2001; 134(8): 663-94.

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

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

Bramer W, GB de Jonge, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. JMLA, 2018; 106(4): 531-41.

Chalmers I. 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, 2003; 589.1: 22-40.

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

Cochrane Handbook, chapter 8, section 8.3.1. https://training.cochrane.org/handbook/current/chapter-08.

Curran-Everett D and Milgrom H. Post-hoc data analysis: benefits and limitations. Current opinion in Allergy and Immunology, 2013; 13(3): 223-4.

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

Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ, 1997; 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. GRADE guidelines 6. Rating the quality of evidence--imprecision. Journal of Clinical Epidemiology, 2011; 64(12): 1283-93.

Hammer GP, du Prel JB, Blettner M. Avoiding bias in observational studies. Dtsch Arztebl Int., 2009; 106(41): 664-8.

Hariton E and Locascio JJ. Randomised controlled trials – the gold standard for effectiveness research. BJOG, 2018; 125(13): 1716.

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

Health Canada. Guidance document for preparing a submission for food health claims. March 2009. Guidance Document for Preparing a Submission for Food Health Claims - Canada.ca

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. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ, 2011; 343: d5928.

Higgins JPT, Thomas J (senior editors). Cochrane Handbook for Systematic Reviews of Interventions Version 6.3 [updated 2022]. The Cochrane Collaboration, 2009. Available from https://training.cochrane.org/handbook/current.

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

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

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

Ministry of Health. 2020. Eating and Activity Guidelines for New Zealand Adults: Updated 2020. Wellington: Ministry of Health.

Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. JAMA, 2006; 295(6): 676-80.

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

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

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

Roberts C, Torgerson DJ. Baseline imbalance in randomised controlled trials. BMJ., 1999. 319(7203): 185.

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

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

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

Sullivan, GM and Feinn, R. Using Effect Size-or Why the P Value Is Not Enough. Journal of Graduate Medical Education, 2012; 4(3): 279–82.

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. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Medicine, 2007; 4(10): e297.

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

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