U. S. Food and Drug Administration
Center for Food Safety and Applied Nutrition
Office of Special Nutritionals
December 22, 1999

Guidance for Industry Significant Scientific Agreement in the Review of Health Claims for Conventional Foods and Dietary Supplements

Comments and suggestions regarding this document should be submitted by February 22, 2000, to Dockets Management Branch (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm 1061, Rockville, MD 20852. All comments should be identified with the docket number 99D-5424.

For questions concerning the content of the document contact Sharon Ross at 202-205-4168.

Additional copies of this guidance document are available upon written request from the Office of Special Nutritionals (HFS-450), Food and Drug Administration, 200 C Street SW, Washington, DC 20204, by calling 202-205-4168, by faxing a request to 202-205-5295, or from the Internet at http://www.cfsan.fda.gov/~dms/guidance.html#lab.

U. S. Department of Health and Human Services
Food and Drug Administration
Center for Food Safety and Applied Nutrition
December 22, 1999

Table of Contents

INTRODUCTORY NOTE

I. BACKGROUND INFORMATION
II. SCIENTIFIC REVIEW OF HEALTH CLAIMS

A. Identifying Data for Review
B. Performing Reliable Measurements
C. Evaluating Individual Studies

D. Evaluating the Totality of the Evidence
E. Assessing Significant Scientific Agreement

III. REFERENCES
Figure 1. Type of Study Evaluated During Health Claim Review
Figure 2. Schema for Assessing Strength and Consistency of Significant Scientific Agreement

Introductory Note

This guidance has been prepared by the Office of Special Nutritionals in the Center for Food Safety and Applied Nutrition at the Food and Drug Administration (FDA), based on the report of the FDA Food Advisory Committee (FAC) Working Group on Significant Scientific Agreement. This guidance represents the agency's current thinking on the meaning of the significant scientific agreement standard in section 403 (r)(3) of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. § 343(r)(3)) and 21 CFR § 101.14(c). It is being issued as level 1 guidance for immediate implementation in accordance with FDA's good guidance practices (62 FR 8961, February 27, 1997). The guidance document does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statute, regulations or both.

This guidance document addresses the significant scientific agreement standard, which FDA uses to evaluate the scientific evidence supporting health claim petitions about the relationship between a food substance and a disease or health-related condition. The guidance document describes the scientific review approach FDA has taken in previous health claim reviews and incorporates the recommendations of the FDA FAC Working Group on Significant Scientific Agreement. This approach is used by FDA scientists in their review of health claims and should be considered as guidance by those compiling health claim petitions. The scientific principles described in this document should also be useful to those designing studies to support health claim petitions.

FDA's determination on significant scientific agreement represents the agency's best judgment as to whether qualified experts would likely agree that the scientific evidence supports the substance/disease relationship that is the subject of a proposed health claim. The significant scientific agreement standard is intended to be a strong standard that provides a high level of confidence in the validity of a substance/disease relationship. Significant scientific agreement means that the validity of the relationship is not likely to be reversed by new and evolving science, although the exact nature of the relationship may need to be refined. Application of the significant scientific agreement standard is intended to be objective, in relying upon a body of sound and relevant scientific data; flexible, in recognizing the variability in the amount and type of data needed to support the validity of different substance/disease relationships; and responsive, in recognizing the need to re-evaluate data over time as research questions and experimental approaches are refined. Significant scientific agreement does not require a consensus or agreement based on unanimous and incontrovertible scientific opinion. However, on the continuum of scientific discovery that extends from emerging evidence to consensus, it represents an area on the continuum that lies closer to the latter than to the former.

Before significant scientific agreement can be assessed, a number of sequential threshold questions are addressed in the review of the scientific evidence:

- Have studies appropriately specified and measured the substance that is the subject of the claim?
- Have studies appropriately specified and measured the disease that is the subject of the claim?
- Are any and all conclusions about the substance/disease relationship based on the totality of publicly

available scientific evidence?

The assessment of significant scientific agreement then derives from the conclusion that there is a sufficient body of sound, relevant scientific evidence that shows consistency across different studies and among different researchers and permits the key determination of whether a change in the dietary intake of the substance will result in a change in a disease endpoint.

The specific topics addressed in this guidance document are: identifying data for review, performing reliable measurements, evaluating individual studies, evaluating the totality of the evidence, and assessing significant scientific agreement. Other aspects of and requirements for the health claim authorization process are described in the Code of Federal Regulations, in 21 CFR § 101.14 and 21 CFR § 101.70.

Major considerations in the scientific review process for health claims are highlighted in bold-face type. For each step in the process, details of the issues that should be considered are provided. Explanatory comment, illustrative discussion points, and examples of application of criteria or requirements, as demonstrated by past health claim authorization reviews, are provided in italics.

Background Information

The Nutrition Labeling and Education Act of 1990 (NLEA) was designed to give consumers more scientifically valid information about the foods they eat (1). Among other provisions, NLEA authorized FDA to allow statements that describe the relationship between a nutrient and a disease or health-related condition to appear in the labeling of foods, including dietary supplements. Such statements about substance/disease relationships are known as "health claims." FDA has defined the term "substance" by regulation as a specific food or component of food. An authorized health claim may be used on both conventional foods and dietary supplements, assuming that the substance in the product and the product itself meet the appropriate standards. Health claims are directed to the general population or designated subgroups (e.g., the elderly) and are intended to assist the consumer in maintaining healthful dietary practices.

When FDA decides whether to authorize a health claim, it evaluates, among other considerations, whether the evidence supporting the relationship that is the subject of the claim meets the significant scientific agreement standard. This standard derives from 21 U.S.C. § 343(r)(3)(B)(i), which provides that FDA shall authorize a health claim to be used on conventional foods if the agency "determines, based on the totality of the publicly available scientific evidence (including evidence from well-designed studies conducted in a manner which is consistent with generally recognized scientific procedures and principles), that there is significant scientific agreement, among experts qualified by scientific training and experience to evaluate such claims, that the claim is supported by such evidence." This scientific standard applies to conventional food health claims by statute; FDA applied the same standard to dietary supplement health claims by regulation. See 21 CFR § 101.14(c).

The NLEA identified 10 substance/disease relationships for initial consideration(1). Of these, significant scientific agreement was determined to exist for eight of the relationships, and health claims describing these relationships on food labels were authorized in 1993. The legislation also permits any interested person to petition FDA to issue a regulation regarding a health claim. Additional health claims have been authorized in response to such petitions.[1]

Since NLEA was enacted, several groups have evaluated the health claim authorization process, including

the interpretation of significant scientific agreement. One of these evaluations was a 2-year Keystone Center dialogue among representatives from academia, industry, consumer groups, and government. The dialogue and resulting report affirmed the principles and approach FDA had been using to authorize health claims(2). The Commission on Dietary Supplement Labels examined the health claim authorization process for dietary supplements and also generally expressed agreement with FDA's approach in its report (3). Following the Keystone dialogue, the FDA FAC convened a number of working groups in 1996 to address issues raised and recommendations made during the dialogue. The FAC Working Group on Significant Scientific Agreement was charged with developing a guide for preparing health claim petitions. In response to the recent decision of the United States Court of Appeals for the District of Columbia Circuit in Pearson v. Shalala, 164 F.3d 650 (D.C. Cir. 1999), which required FDA to clarify the meaning of significant scientific agreement, the focus of the FAC Working Group shifted to the scientific review of data for health claims and the interpretation of the significant scientific agreement standard. The final report of the FAC Working Group on Significant Scientific Agreement, entitled "Interpretation of Significant Scientific Agreement in the Review of Health Claims," was made public during the FAC meeting of June 24-25, 1999. (See http://vm.cfsan.fda.gov/~dms/facssa.html for a copy of the Working Group's report.) Following additional comment by the FAC, FDA adopted the recommendations proposed by the Working Group on Significant Scientific Agreement. This guidance document is based on the FAC Working Group report but has been expanded and edited to clarify and more fully explain some topics. The guidance represents the agency's current thinking on the meaning of significant scientific agreement in 21 U.S.C. § 343(r)(3)(B)(i) and 21 CFR § 101.14(c).

Footnote:

1. In 1997, Congress enacted the Food and Drug Administration Modernization Act, which established an alternative authorization procedure for health claims based on authoritative statements from certain federal scientific bodies or from the National Academy of Sciences. As of December 1999, one health claim had been authorized under this alternative procedure. This guidance document does not address that alternative procedure.

Scientific Review of Health Claims

The scientific review process FDA uses to evaluate health claims is comprehensive and focuses first on review of individual studies. After identifying relevant, good quality studies and assessing their strengths and weaknesses, the agency conducts a more comprehensive review based on the body of evidence as a whole. Considerations in the scientific review of health claims are detailed below.

The standard of scientific validity for a health claim includes two components: 1) that the totality of the publicly available evidence supports the substance/disease relationship that is the subject of the claim, and 2) that there is significant scientific agreement among qualified experts that the relationship is valid.

FDA's evaluation of the evidence supporting a health claim is based on the totality of publicly available data. Because of the limitations of the various research methods that can be used to study substance/disease relationships, it is not possible to specify the type or number of studies needed to support a health claim. In addition, each relationship involves a unique set of confounders (see discussion below) and measurement issues.

Sound, relevant science in research design and measurement -- to ensure that research, in fact,

provides the answers to the questions that need to be addressed concerning the relationship -- drives the decision to authorize health claims, not the specific type or number of studies. This point is illustrated graphically in Figure 1, which shows the number and nature of the human studies evaluated in determining the validity of certain of the initial health claims evaluated during the 1990-1992 review and claims for which petitions were submitted. The number and types of studies considered varied greatly among authorized claims.

In addition to limitations imposed by available research methods, another limitation frequently encountered is the dependence on publicly available data derived from studies that were not specifically designed or conducted for the purpose of supporting a health claim. Thus, in the agency's review of health claims, the usefulness, relevance, and generalizability of such studies to the health claim under consideration are carefully evaluated, especially in terms of specification and measurement of the substance and disease whose relationship is the subject of the claim.

A. Identifying Data for Review

The first step in preparing or reviewing a health claim petition is to identify all relevant studies.

The types of studies considered in a health claim review include human studies and frequently also include "pre-clinical" evidence, e.g., in vitro laboratory investigations and other mechanistic studies. Studies of humans can be divided into two types: interventional studies and observational studies.

In an interventional study, the investigator controls whether the subjects receive an exposure or an intervention whereas in an observational study, the investigator does not have control over the exposure or the intervention. In general, interventional studies provide the strongest evidence for an effect.

Regardless of the inherent strengths and weaknesses of a study design, the overall quality and relevance of each individual study is paramount in assessing its contribution to the weight of the evidence for the proposed substance/disease relationship.

- Interventional studies

The "gold standard" of interventional studies is the randomized controlled clinical trial.

In a randomized controlled trial, subjects similar to each other are randomly assigned either to receive the intervention or not to receive the intervention. As a result, subjects who are most likely to have a favorable outcome independent of any intervention are not preferentially selected to receive the intervention being studied (selection bias). Bias may be further reduced if the researcher who assesses the outcome does not know which subjects received the intervention (blinding). Randomized controlled clinical trials are not an absolute requirement to demonstrate significant scientific agreement in all cases, but are considered the most persuasive and given the most weight. A single large, well-conducted and controlled clinical trial could provide sufficient evidence to establish a substance/disease relationship, provided that there is a supporting body of evidence from observational or mechanistic studies.

Interventional studies for foods may differ from those for drugs. Unlike drug studies, food interventional trials may have additional confounders secondary to using a food substance as

the intervention (see discussion below). In addition, it may not be possible to use a placebo control group for food studies, and subjects in such studies may not be blinded to the intervention. As a result of the greater likelihood for confounders and bias, interventional studies with foods may generate data that have less certainty than data from drug interventional studies.

Although interventional studies are the most reliable category of studies for determining cause-and-effect relationships, generalizing from selected populations often presents serious problems in the interpretation of such studies. Furthermore, in some cases, such as with cancers of different sites, interventional dietary studies are not feasible because diseases with lower frequency of occurrence, such as rare forms of cancer, require very large study samples to detect an effect. Moreover, there frequently are long delays from dietary exposure to onset of disease, often 20 to 30 years. Therefore, the scientific evidence supporting a substance/disease relationship may have to be derived wholly or in part from observational studies.

- Observational studies

There is no universally valid method for weighing categories of observational studies. However, in general, observational studies include, in descending order of persuasiveness, cohort (longitudinal) studies, case-control studies, cross-sectional studies, uncontrolled case series or cohort studies, time-series studies, ecological or cross-population studies, descriptive epidemiology, and case reports.

Observational studies may be prospective or retrospective. In prospective studies, investigators recruit subjects and observe them prior to the occurrence of the outcome. In retrospective studies, investigators review the records of subjects and interview subjects after the outcome has occurred. Retrospective studies are usually considered to be more vulnerable to recall bias (error that occurs when subjects are asked to remember past behaviors) and measurement error but are less likely to suffer from the subject selection bias that may occur in prospective studies.

- *Cohort studies compare the outcome of subjects who have received a specific exposure with the outcome of subjects who have not received that exposure.*
- *In case-control studies, subjects with the disease are compared to subjects who do not have the disease (control group). Subjects are enrolled based on their outcome rather than based on their exposure.*
- *In cross-sectional studies, at a single point in time the number of individuals with a disease who have received a specific exposure is compared to the number of individuals without the disease who did not receive the exposure.*
- *Uncontrolled case series studies depict outcomes in a group without comparing to a control group.*
- *Time-series studies compare outcomes during different time periods, e.g., whether the rate of occurrence of a particular outcome during one five-year period changed during a subsequent five-year period.*
- *In ecological studies, the rate of a disease is compared across different populations.*

Investigators seek to identify population traits that may cause the disease.

- *Descriptive epidemiology refers to study designs that assess parameters related to the frequency and distribution of disease in a population, such as the leading cause of death.*
- *Case reports describe observations of a single subject or a small number of subjects.*

A common weakness of observational studies is the limited ability to ascertain the actual food or nutrient intake for the population studied. Observational data are also generally restricted to identifying associations between food substances and health outcomes, and often do not provide a sufficient basis for determining whether a substance/disease association reflects a causal rather than a coincidental relationship.

- Research synthesis studies

  "Research synthesis" studies, including meta-analyses, may be useful as supporting evidence for a health claim, but any role beyond this function is as yet unresolved.

  The appropriateness of research synthesis studies to establish substance/disease relationships is not known. This is especially true when observational data are entered into meta-analyses. Discussions on the topic have been published (4-7), and there are on-going efforts to identify criteria and critical factors to consider in both conducting and using such analyses, but standardization of this methodology is still emerging. Therefore, in general, such analyses serve as supporting evidence rather than as primary evidence. To date, while meta-analyses have been reviewed as part of the health claim authorization process, no health claims have been authorized on the basis of meta-analysis studies alone.

- Animal and in vitro studies

  Although human studies are weighted most heavily in the evaluation of evidence on a substance/disease relationship, data from animal model and in vitro (laboratory) studies also can be used to support a substance/disease relationship.

  Lacking any data from human studies, animal and in vitro studies alone would not adequately support a health claim. Although both types of studies permit greater control over variables, such as diet and genetics, and permit more aggressive intervention, each suffers from the uncertainties of extrapolating to physiological effects in humans. However, these studies can be useful in providing information on the mechanism of action and specificity of a food substance and the process that causes a disease or health-related condition. Animal and in vitro studies should be considered when there are problems designing interventional studies or in the absence of an appropriate biomarker. If such studies are used, they are subjected to the same kind of assessment as the human studies. In the case of animal studies, the consistency of the demonstrated association between a substance and the disease or health-related condition is important when considering whether evidence from such studies supports a health claim. Thus, the strongest animal evidence would be based on data derived from studies on appropriate animal models, on data that have been reproduced in different laboratories, and on data that give a statistically significant dose-response relationship.

B. Performing Reliable Measurements

Appropriate measurement, of both the substance and the disease or health-related condition, is a key factor in the review of data for health claims.

Assessing the effects of diet on human health is limited by a variety of measurement issues: the use of biomarkers, the difficulty of identifying and measuring the food substance that provides the effect, the difficulty of accurately measuring dietary intake, and the difficulty of distinguishing the effects of diet on a disease from those of other variables, such as weight change, physical activity, or environmental factors.

- Biomarkers

  Because a number of the diseases associated with dietary factors are diseases that develop over a period of many years (chronic diseases), a person may not show outward signs or symptoms of a disease at a particular stage of the illness even though that person has the disease. For example, individuals may have deposits of fat and other material accumulating in the arteries to their hearts (atherosclerotic coronary heart disease) and not experience any symptoms until years later when they suffer a heart attack. Therefore, scientists seek to identify "biomarkers" (intermediate or surrogate endpoint markers) for the presence or risk of disease.

  A biomarker is a measurement of a variable related to a disease that may serve as an indicator or predictor of that disease. Biomarkers are parameters from which the presence or risk of a disease can be inferred, rather than being a measure of the disease itself. In conducting a health claim review, FDA does not rely on a change in a biomarker as a measurement of the effect of a dietary factor on a disease unless there is evidence that altering the parameter can affect the risk of developing that disease or health-related condition. This is the case for serum cholesterol in that high levels are generally accepted as a predictor of risk for coronary heart disease, and there is evidence that decreasing high serum cholesterol can decrease that risk. Therefore, the evaluation of whether decreasing the intake of dietary fat reduces the risk of developing heart disease took into account many studies that assessed changes in serum cholesterol, specifically LDL-cholesterol, rather than the development of heart disease per se. For the existing authorized health claims, acceptable biomarkers are LDL-cholesterol levels for coronary heart disease, measures of bone mass for osteoporosis, and measures of blood pressure for hypertension.

- Identifying and measuring the food substance

  The measurement of a food substance centers on the following questions: 1) What was measured? and 2) How does the measured substance relate to the substance that is the subject of the health claim?

  Studies that examine dietary components often focus on the intake of the substance of interest as part of a food or a total diet, or may infer intake as part of post-hoc evaluations of the data. Therefore, isolating the effect of the substance can be a critical consideration in authorizing a health claim. Common difficulties involve separating the effect of the food substance from the food itself, or the use of measures that reflect heterogeneous or poorly defined food substances. Without evidence that the substance, rather than the overall diet or specific foods in the diet, is responsible for the benefit, the linkage between the substance and the disease cannot be established.

  FDA applied this principle during evaluations of the initial 10 substance/disease relationships in 1990-1992. In the case of claims related to omega-3 fatty acids, fiber, and antioxidant vitamins, there was considerable measurement overlap between the food containing the

substance and the substance itself, or there were concomitant changes in other dietary components. Fiber was poorly defined and/or a heterogeneous mixture as measured in research available at the time of the initial health claim review. For example, as noted during the health claim review for fiber and heart disease, the objective of the protocols of many studies was to evaluate the effectiveness of relatively large amounts of a single type of food or fiber source rich in soluble fiber (e.g., baked beans), rather than to examine total soluble dietary fiber intakes or to specifically identify the chemical and physical characteristics of soluble fiber that are most effective in lowering blood cholesterol levels. Thus, the effects could not be attributed to the fiber. Moreover, in some studies large amounts of foods (e.g., 1-2 cups of baked beans) were added to diets; these dietary changes were often accompanied by lower calorie intakes with resultant weight loss, which has an independent impact on the risk of developing heart disease.

Measurement issues generally focus on substances in food, but the same principles apply when the substance of interest is itself a food. While a single food can be the subject of a health claim, existing experience is that the subject is more likely to be a group of foods, such as fruits, vegetables, and grains, which have been associated with a reduced risk of heart disease and of cancer. This identification, and consequently measurement, of a food group is, in turn, most likely to occur because it is not possible to identify and, therefore, measure a particular component of these foods that is responsible for the benefit. Nonetheless, in theory, it is possible that a unique combination of nutrients or other substances in a single food could be the subject of a health claim. To date, this has not occurred.

- Assessment of dietary intake

In determining whether a substance that is the subject of a claim has been measured appropriately, it is important to evaluate critically the method of assessment of dietary intake. Each method has its strengths and weaknesses. No one method is adequate for every purpose.

Dietary intake assessment methods include food records, 24-hour recalls, and diet histories. Food records are based on the premise that food weights provide an accurate estimation of food intake. Subjects weigh the foods they consume and record those values. The 24-hour recall method requires that subjects describe which foods and how much of each food they consumed during the prior 24-hour period. Diet histories use questionnaires or interviewers to estimate the typical diet of subjects over a certain period of time. For a more detailed description of these methods and their strengths and weaknesses, see Diet and Health (8). Some common problems that weaken confidence in the assessment of dietary intake may be noted. For example, a single 24-hour recall is generally regarded as an inadequate method for assessing the usual intake of a nutrient or other food substance by an individual, although it may be useful for assessing mean intake of a group. A diet history taken by a food frequency questionnaire that contains a limited number of items is inadequate for assessing intake of a specific nutrient if the major food sources of the nutrient in the population studied are not included in the questionnaire. Finally, accurate estimation of the intake of a nutrient or other food substance derived from any type of intake assessment is also dependent on the availability of valid and complete food composition databases for the nutrient or other substance of interest.

- Distinguishing the effects of diet from other variables

Scientific studies provide the means to identify which effects on a disease or health-related condition result from the consumption of a particular food substance and which effects are the products of other

factors. Evaluating the conclusions of a study requires an assessment of both the design and conduct of the study, as well as the methods used to interpret the data obtained from the study. Appropriate control of potential confounding factors, by eliminating as many as possible in interventional studies and by adjusting for them with appropriate data analysis techniques in observational studies, is needed if studies are to contribute substantively to the weight of evidence in support of a substance/disease relationship.

C. Evaluating Individual Studies

The evaluation of study design, protocol, measurement, and statistical issues for individual studies serves as the starting point from which FDA determines the overall strengths and weaknesses of the data and assesses the weight of the evidence.

FDA's review of individual studies on substance/disease relationships generally follows the approaches outlined in the Guide to Clinical Preventive Services (9) and Diet and Health (8).

The persuasiveness of a study depends on the quality of the study.

> *Evaluation of the quality of individual studies on substance/disease relationships begins with a consideration of the inherent strengths and weaknesses of various study designs. The three most important measures of the quality of a study are design, conduct, and analysis and interpretation.*

- Bias and confounders

  Certain study designs tend to be more persuasive because they are less subject to bias and measurement error. As noted earlier, retrospective studies are usually considered to be more vulnerable to recall bias and measurement error but are less likely to suffer from the subject selection bias that may occur in prospective studies. Different degrees of persuasiveness may also be assigned within classes of studies, depending on the particular assessments made. For example, case-control studies in which higher or lower serum levels of a nutrient or metabolite are found in cases versus controls will generally be less persuasive in establishing a substance/disease relationship than similar studies that assess an antecedent behavior (such as dietary intake), despite the potential for recall bias, because such studies cannot distinguish whether the high or low serum level of the nutrient was a contributing cause or a consequence of the disease.

> *The susceptibility of research data to bias and confounding depends on several factors, including the methods used to choose subjects and to measure outcomes, the use of a comparison (control) group, and whether the study was conducted retrospectively or prospectively. Confounders are factors that are associated with the disease in question and the intervention, and that prevent the measured outcome from being attributed unequivocally to the intervention.*
>
> *Several aspects of substance/disease relationships may give rise to confounders. Foods are rarely composed of a simple mixture of chemical constituents. The addition of a nutrient to a diet, or an increase in total daily intake of that nutrient, may have unintended effects. The added nutrient may displace other nutrients in the diet. Therefore, it may be difficult to ascertain whether the health outcome is the result of the added nutrient or the related changes*

on the original diet. For example, weight loss was a confounder in a number of studies used to support a claim that lowering of dietary saturated fat intake and resultant decreases in serum LDL-cholesterol led to a reduced risk of coronary heart disease. Diets low in fat can result in a lower calorie intake and, in turn, weight loss. Since weight loss per se can reduce levels of LDL-cholesterol, the benefit in those studies could not be attributed to the lack of the food substance (saturated fat), but may have been related to lower calorie intake. Nonetheless, sufficient studies that did control for such related factors were available and there was adequate evidence to establish a relationship between diets low in saturated fat and cholesterol and reduced risk of heart disease. Other potential confounders include variability in the quantity or quality of the food substance being administered.

- Quality assessment criteria

  Criteria that are considered in assessing the quality of individual studies of substance/disease relationships include the following:

  - Adequacy and clarity of the design

    Were the questions to be answered by the study clearly described at the outset?

    Was the methodology used in the study clearly described and appropriate for answering the questions posed by the study?

    Was the duration of the study intervention or follow-up period sufficient to detect an effect on the outcome of interest?

    Were potential confounding factors identified, assessed, and/or controlled?

    Was subject attrition (subjects leaving the study before the study is completed) assessed, explained, and reasonable?

  - Population studied

    Was the sample size large enough to provide sufficient statistical power to detect a significant effect? (If the study is underpowered, it may be impossible to conclude that the absence of an effect is not due to chance.)

    Was the study population representative (for factors such as age, gender distribution, race, socioeconomic status, geographic location, family history, health status, and motivation) of the population to which the health claim will be targeted?

    Were criteria for inclusion and exclusion of study subjects clearly stated and appropriate?

    Were recruitment procedures that minimized selection bias used?

    For controlled interventions, were subjects randomized? If matching was employed to assign the subjects to control and treatment groups, were appropriate demographic characteristics and other variables used for the matching? Was randomization successful in producing similar control and intervention groups?

- Assessment of intervention or exposure and outcomes

  Were analytical methodology and quality control procedures to assess dietary intake adequate?

  Was the dietary intervention or exposure well defined and appropriately measured? (See discussion above.)

  For intervention studies, was an appropriate level of intake (i.e., the level hypothesized to be effective) for the food substance of interest planned, monitored, and achieved?

  Were the background diets to which the test substance was added, or the control and interventional diets, adequately described, measured, and suitable?

  Was a "lead-in" period employed for dietary interventions? (Because changes in the diet may induce compensatory metabolic changes, the effect of an intervention should be measured after stabilization has occurred, i.e., a lead-in period.)

  In studies with cross-over designs, was there an appropriate "wash-out" period (period during which subjects do not receive an intervention) between dietary treatments? (Lack of a sufficient wash-out period between interventions may lead to confusion as to which intervention produced the health outcome.)

  Were the form and setting of the intervention representative of the "real world?"

  Were other possible concurrent changes in diet or health-related behavior (weight loss, exercise, alcohol intake, smoking cessation) during the study that could account for the outcome identified, assessed, and/or controlled?

  Were the disease outcomes well defined and appropriately measured? If biomarkers (intermediate or surrogate endpoint markers) were measured, has their relevance to disease outcomes been validated?

  Were efforts made to detect harmful as well as beneficial effects? (For example, increasing the consumption of some food substances may increase the risk of a chronic disease, and extracting or concentrating some food substances may render them injurious to health.)

- Statistical methods

  Were appropriate statistical analyses applied to the data?

  Was "statistical significance" interpreted appropriately? (For example, differences that are not statistically significant should be described as not demonstrating a difference rather than as showing a trend.)

  Were relative and absolute effects distinguished?

- Summary of the evidence

As part of the review process, FDA creates a summary of the scientific evidence to help organize and guide its comprehensive review. FDA recommends that health claim petitions include a summary of the evidence describing the individual studies in table form. Such summaries help speed agency review of the petition.

D. Evaluating the Totality of the Evidence

Evaluating the totality of the evidence means evaluating whether it permits the key determination of whether a change in the dietary intake of the substance will result in a change in a disease endpoint.

After identifying relevant, good quality studies and assessing and summarizing their strengths and weaknesses, FDA conducts a more comprehensive review based on the body of evidence as a whole. Petitioners should be sure that the conclusions the petition draws regarding the association between nutritional exposures or interventions and outcomes are objectively based on the totality of the evidence, and that interpretations are limited to the research conducted, without inappropriate extrapolations beyond the available evidence.

A classic set of reviews that demonstrate an appropriate process for evaluating substance/disease relationships is the work conducted by The Task Force on The Evidence Relating Six Dietary Factors to the Nation's Health (10). Its approach incorporated the standard principle that the strength of evidence associating a nutritional exposure with a health outcome depends not only on the quality of the individual studies but on the overall "grade" or assessment of the evidence taken together, the number of studies, consistency of results, and the magnitude of effects.

- Determining the strength of the substance/disease association

  The strength of evidence that exposure to a particular food substance is associated with a health outcome depends on several factors.

  - The first consideration in judging the body of evidence is determining whether most of the evidence is derived from more persuasive classes of study designs.

    The design category and the quality of the research methodology should be considered together. Various coding and scoring schemes have been devised to systematize this process. The U.S. Preventive Services Task Force's grading system assigns a letter code to rate the quality of the evidence (9). Other groups have developed systems that score a study quantitatively, assigning points for different aspects of design quality and performance (11). However, although both study design codes and quantitative scores are appropriate for rating individual studies, they do not adequately describe the evidence as a whole. For example, these methods do not capture the number of studies or consistency of findings. At present, a universally applicable system for evaluation of the evidence as a whole is not available.

  - Another contribution to the strength of the evidence is the number of studies in support of the association.

    The number of studies required to be persuasive is often inversely related to the overall class of evidence available. Simply counting the studies with positive results without regard for their

individual quality is an inadequate approach to assessing the overall strength of the evidence.

- Consistency of results across different settings and types of populations also bolsters the strength of an association.

  Conflicting results do not disprove an association (because elements of the study design may account for the lack of an effect in negative studies) but do tend to weaken confidence in the strength of the association. In general, the greater the consistency, the more likely the significant scientific agreement standard will be met. However, repetition of a poorly designed study does not add to the consistency or quality of the evidence.

- Finally, if the magnitude of the effect is large, yielding strong statistical significance and narrow confidence intervals, evidence of an association is bolstered and the association is more likely to have clinical significance.

- Determining the strength of the substance/disease relationship (inferring that a causal relationship exists)

  Evidence of an association does not, however, prove cause and effect. An association of variables only indicates that they occur together but not that one causes the other. Therefore, another step in the process of a health claim review is to determine the strength of the evidence for a causal relationship.

A causal relationship exists when data show that the consumption of a substance increases or decreases the probability of developing or not developing a particular disease or health-related condition. Causality can be best established by interventional data, particularly from randomized, controlled clinical trials, that show that altering the intake of an appropriately identified and measured substance results in a change in a valid measure of a disease or health-related condition. In the absence of such data, a causal relationship may be inferred based on observational and mechanistic data through strength of association, consistency of association, independence of association, dose-response relationship, temporal relationship, effect of dechallenge, specificity, and explanation of a pathogenic mechanism or a protective effect against such a mechanism (biological plausibility). Although these features strengthen the claim that a substance contributes to a certain health outcome, they do not prove that eating more or less of the substance will produce a clinically meaningful outcome. In many cases (for example, if the intake of the substance has not been or cannot be assessed adequately in available observational studies because it has not been commonly consumed or its intake cannot be assessed independently of other substances), controlled clinical trials are necessary to establish the validity of a substance/disease relationship.

- *Strength of association is sometimes described as relative risk. Relative risk is the ratio between the rate of disease for subjects exposed to the substance and the rate for subjects not exposed. The larger the relative risk, the more likely that ingesting the substance is causally related to the health outcome.*

- *Consistency of association means that the same association is found across several studies and among various population groups.*

- *Independence of association refers to the extent to which the association relates to the exposure or intervention being studied versus the extent to which the association relates to a variable other than the exposure or intervention.*

- *Dose-response relationship means that greater effects occur with greater exposures to the*

substance.

- *Temporal relationship means that the exposure consistently precedes the outcome.*
- *Effect of dechallenge means that subjects from whom the intervention has been withdrawn demonstrate a reversal of the associated outcome.*
- *Specificity means the degree to which the substance is associated only with the disease in question. The more specific an association, the more likely the association is causal. However, lack of specificity may not be a critical factor in the assessment of substance/disease relationships because many etiological agents cause more than one disease, and many diseases have multifactorial causes.*
- *Biological plausibility means that there is a biological explanation for the causal relationship. Although biological plausibility is not necessary to infer causality, it enhances the case.*

- Determining the weight of the evidence as a whole

In assessing whether the totality of the evidence supports the substance/disease relationship that is the subject of the claim, FDA asks two questions:

1. Does the evidence in support of the substance/disease relationship outweigh the evidence against it? In considering this question, appropriate weight should be given to studies that are more persuasive because of the quality of the study design, conduct, and analysis.
2. Is the available body of evidence sufficient to permit the conclusion that a change in the dietary intake of the substance will result in a change in the disease endpoint?

E. Assessing Significant Scientific Agreement

Assessing significant scientific agreement relies on judging the extent of agreement among qualified experts.

Significant scientific agreement refers to the extent of agreement among qualified experts in the field. In the process of scientific discovery, significant scientific agreement occurs well after the stage of emerging science, where data and information permit an inference, but before the point of unanimous agreement within the relevant scientific community that the inference is valid. The significant scientific agreement standard is met when the validity of the relationship is not likely to be reversed by new and evolving science, although the exact nature of the relationship may need to be refined over time. Significant scientific agreement can be achieved when the validity of a substance/disease relationship is supported by the conclusions of federal government scientific bodies; conclusions of independent, expert bodies may also be relevant. When such conclusions are not available (for instance, if the data supporting a proposed health claim are relatively new and have not yet been reviewed by an independent, expert panel or body), a compelling and relevant body of evidence may nonetheless cause the agency to conclude that significant scientific agreement exists.

Although significant scientific agreement is not consensus in the sense of unanimity, it represents considerably more than an initial body of emerging evidence. Because each situation may differ with the nature of the claimed substance/disease relationship, it is

necessary to consider both the extent of agreement and the nature of the disagreement on a case-by-case basis. If scientific agreement were to be assessed under arbitrary quantitative or rigidly defined criteria, the resulting inflexibility could cause some valid claims to be disallowed where the disagreement, while present, is not persuasive.

In order for qualified experts to reach an informed opinion regarding the claim, the data and information that pertain to the claim must be available to the relevant scientific community.

The usual mechanism to show that the evidence is available to qualified experts is that the data and information are published in peer-reviewed scientific journals. Abstracts generally provide insufficient information for review; however, not all the data need be published. FDA reviews information that is not published as long as that information is placed in the public domain at the time the agency takes action on a health claim petition. The value of an expert's opinion will be limited if he/she did not have access to all the evidence.

- Significant scientific agreement depends on the strength and consistency of the evidence.

  Significant scientific agreement cannot be reached without a strong, relevant, and consistent body of evidence on which experts in the field may base a conclusion that a substance/disease relationship exists. There is considerable potential for incorrect conclusions if only preliminary evidence (emerging science) is available for review.

  This is best illustrated by the body of evidence for the association between beta-carotene and cancer risk. At the time of FDA's health claim review, no results from relevant clinical trials had been reported. However, human epidemiological studies were available, as well as laboratory data for mechanistic theories on how beta-carotene might provide a risk reduction effect. While there was strong evidence that high intakes of fruits and vegetables rich in carotenoids were associated with a reduced risk of developing cancer, it was unclear whether the component(s) of fruits and vegetables responsible for reducing the effect were beta-carotene, other carotenoids, or some other compound(s). However, animal studies strongly pointed to a positive effect of beta-carotene in lowering the frequency and severity of experimental cancer induced in animals. The review concluded, nonetheless, that existing evidence was inconclusive and significant scientific agreement did not exist; the animal studies could not be applied directly to humans because the type and amount of carcinogen exposure in the experimental conditions were not similar to human exposure. Subsequently, the decision was further supported when a randomized, controlled trial in Finland tested the ability of antioxidant vitamins, including beta-carotene, to prevent the development of lung cancer in high-risk Finnish men with a history of smoking (12). The unexpected outcome was a significant increase in the rate of lung cancer among the beta-carotene supplemented group.

Figure 2 provides a graphical representation of the interplay of considerations that contribute to determining whether the significant scientific agreement standard for a substance/disease relationship has been met. It illustrates the manner in which evaluations of the various types and amounts of data that may exist for a substance/disease relationship are combined to assess the overall strength and consistency of the scientific evidence. The schema also demonstrates that the significant scientific agreement standard is one that is objective, flexible, and responsive by illustrating the variety of combinations of data from different types of good quality studies that may give rise to a body of evidence sufficient to establish the validity of a substance/disease relationship.

In determining whether there is significant scientific agreement, FDA takes into account the viewpoints of

qualified experts outside the agency, if evaluations by such experts have been conducted and are publicly available. For example, FDA will take into account:

- *review publications that critically summarize data and information in the secondary scientific literature;*
- *documentation of the opinion of an "expert panel" that is specifically convened for this purpose by a credible, independent body;*
- *the opinion or recommendation of a federal government scientific body such as the National Institutes of Health (NIH) or the Centers for Disease Control and Prevention (CDC); or the National Academy of Sciences (NAS); or an independent, expert body such as the Committee on Nutrition of the American Academy of Pediatrics (AAP), the American Heart Association (AHA), American Cancer Society (ACS), or task forces or other groups assembled by the National Institutes of Health (NIH).*

FDA accords the greatest weight to the conclusions of federal government scientific bodies, especially when the evidence for the validity of a substance/disease relationship has been judged by such a body to be sufficient to justify dietary recommendations to the public. Although reviews by individual outside experts are considered in assessing significant scientific agreement, evidence from such reviews alone would not necessarily support a conclusion that the standard has been met, especially if the conclusions of such reviews were not supported by available assessments of the same body of evidence from federal scientific bodies, expert panels or independent, expert bodies. Reviews by outside experts or expert panels are most useful when there is a reasonable basis to conclude that they represent the larger group of qualified experts in the field. Most importantly, the relevance of an outside expert review depends on whether the evidence examined applies to the claim in terms of considerations such as specification and measurement of the substance and the disease or health-related condition.

References

1. Public Law 101-553, 104 Stat. 2353 codified at 21 USC § 343 (1994). Nutrition Labeling and Education Act. November 8, 1990.

2. The Keystone National Policy Dialogue on Food, Nutrition, and Health: Final Report. Keystone, CO: Keystone Press, 1996.

3. Commission on Dietary Supplement Labels. Commission of Dietary Supplement Labels Report to the President, Congress, and the Secretary of the Department of Health and Human Services. Washington, DC: Office of Disease Prevention and Health Promotion, DHHS, 1997.

4. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med 1987;316:450-455.

5. Sacks HS, Berrier J, Reitman D, Pagano D, Chalmers TC. Meta-analysis of randomized controlled trials: an update. In: Balder WC, Mosteller F, eds. Medical Uses of Statistics, 2nd ed. pp 427-442. Boston, MA: NEJM Books, 1992.

6. Sacks HS. Meta-analyses of clinical trials. In: Perman JA, Rey J, eds. Clinical Trials in Infant Nutrition, Nestle Nutrition Workshop Series, Vol 40, pp 85-99. Philadelphia, PA: Vevey/Lippincott-Raven Publishers, 1998.

7. Hasselblad V, Mosteller F, Littenberg B, Chalmers TC, Hunick MG, Turner JA, et al. A survey of current problems in meta-analysis. Discussion from the Agency for Health Care Policy and Research Inter-PORT Work Group on Literature Review/Meta-Analysis. Med Care 1995;33:202-220.

8. National Research Council. Diet and Health: Implications for Reducing Chronic Disease Risk. Washington, DC: National Academy Press, 1989.

9. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. Report of the US Preventive Services Task Force: Guide to Clinical Preventive Services. 2nd ed. Washington, DC: Office of Public Health and Science, April, 1989.

10. Ahrens EH, Connor WE, eds. Symposium: Report of the Task Force on the Evidence Relating Six Dietary Factors to the Nation's Health. Am J Clin Nutr 1979;23(suppl):2621-2748.

11. Mohar D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: current issues and future directions. Int J Technol Assess Health Care 1996;12:125-208.

12. Albanes D, Heinonen OP, Huttunen JK, Taylor PR, Virtamo J, Edwards BK, Haapakoski J, Rautalahti M, Hartman AM, Palmgren J, et al. Effects of alpha-tocopherol and beta-carotene supplements on cancer incidence in the Alpha-Tocopherol Beta-Carotene Cancer Prevention Study. Am J Clin Nutr 1995;62(6 Suppl):1427S-1430S.

Figure 1 Type of Study Evaluated During Health Claim Review

Relationships indicated in italics were the subjects of health claim petitions; other relationships were evaluated during the NLEA health claim review, 1990-1992. Relationships with * were evaluated but not authorized. Neither the type nor number of studies reviewed determines whether a health claim can be authorized.

Abbreviations used in Relationship legend: Na/Hypert = Sodium and Hypertension; Fat/Ca = Dietary Fat and Cancer; Fat/CVD = Saturated Fat and Cholesterol and Cardiovascular Disease; Omg3/CVD = Omega-3 Fatty Acids and Cardiovascular Disease; Zinc/Immun = Zinc and Immune Deficiency in the Elderly; VitE/Ca = Vitamin E and Cancer; VitC/Ca = Vitamin C and Cancer; Carotn/Ca = Carotenoids and Cancer; Fol/NTD = Folate and Neural Tube Defects; FiberFood/CVD = Fruits/Vegetables/Grain and Cardiovascular Disease; FiberFood/Ca = Fruits/Vegetables/Grain and Cancer; Cal/Osteoporosis = Calcium and Osteoporosis; OatFiber/CHD = Beta-glucan Soluble Fiber from Oats and Coronary Heart Disease; Psyllium/CHD = Soluble Fiber from Psyllium Husk and Coronary Heart Disease.

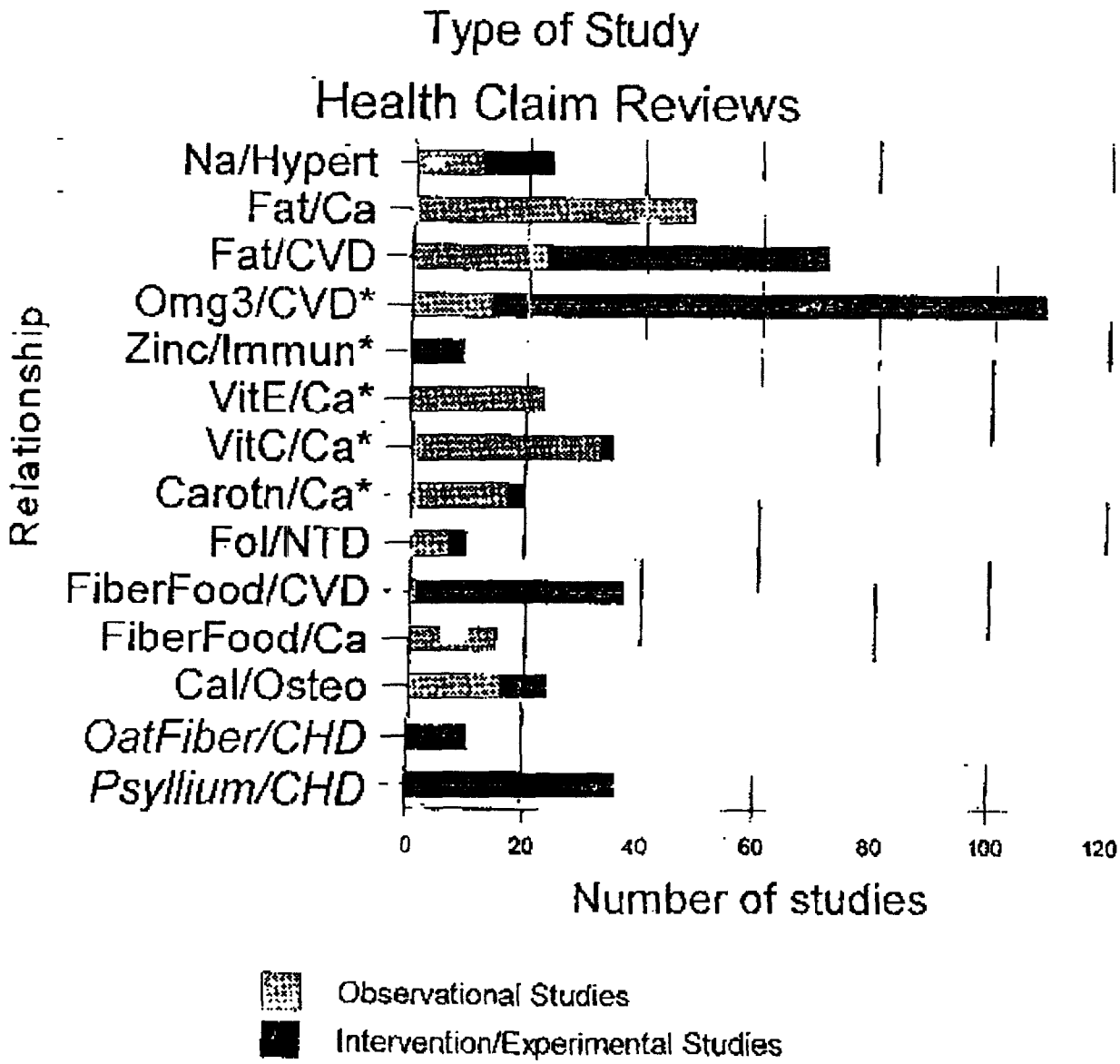
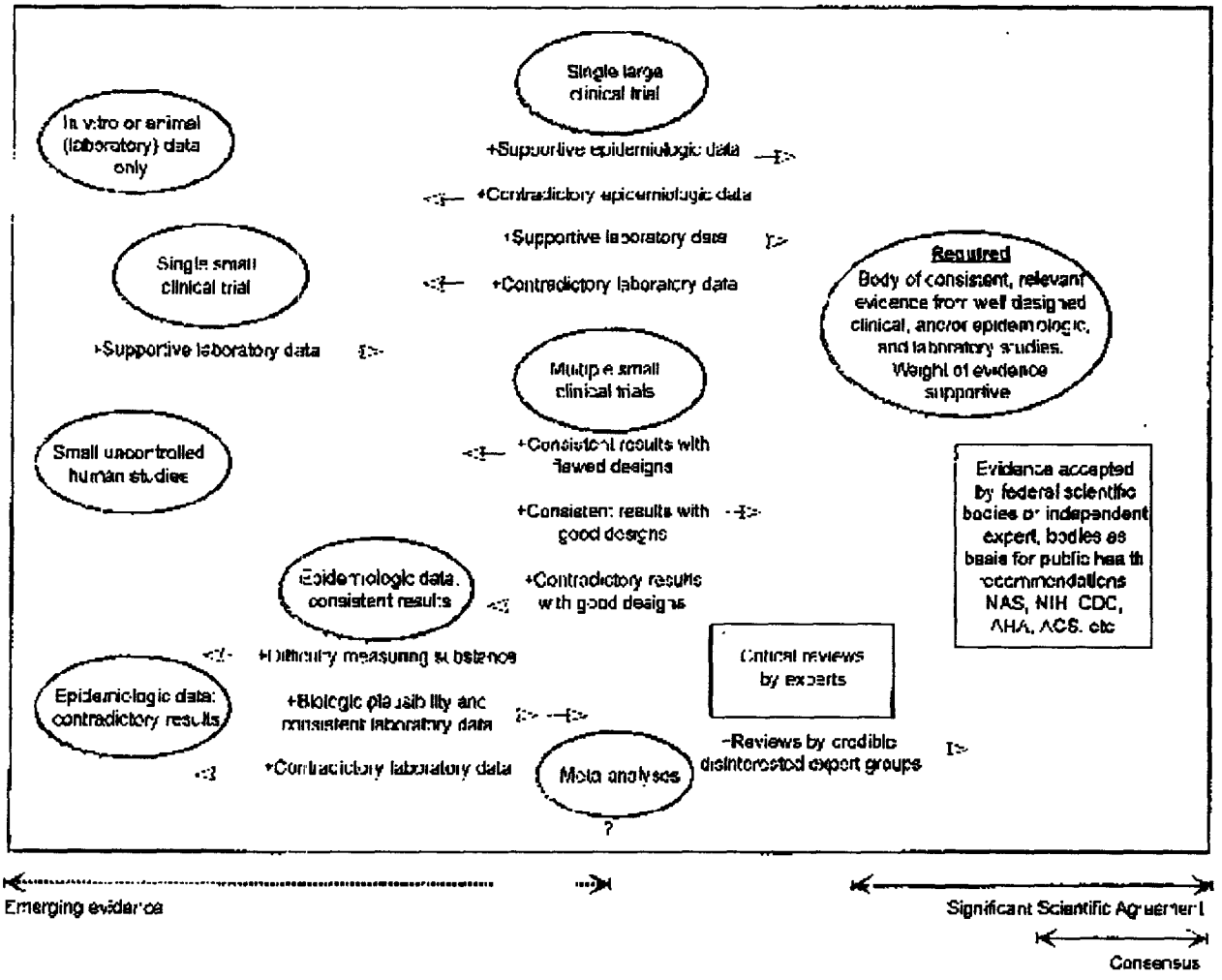


Figure 2. Schema for Assessing Strength and Consistency of Scientific Evidence Leading to Significant Scientific Agreement



This document was issued on December 22, 1999.
For more recent information on Dietary Supplements
See http://www.cfsan.fda.gov/~dms/supplmnt.html

Dietary Supplements

Foods Home | FDA Home | Search/Subject Index | Disclaimers & Privacy Policy

Hypertext updated by ear/kwg 2000-NOV-22