Guidance: Interim Evidence-based Ranking System for Scientific Data

This interim ranking system provides criteria to rank scientific evidence relevant to substance/disease relationships that are the subject of qualified health claims. It outlines the major concepts the Task Force thinks the agency should consider in guiding the scientific evaluation.

The primary purpose of this attachment is to provide petitioners with a description of the major points the Task Force thinks the agency should consider in evaluating supporting scientific data.

What is an Evidence-based Rating System?

An evidence-based rating system is a science-based systematic evaluation of the strength of the evidence behind a statement. In the case of health claims, it would rate the strength of the evidence behind a proposed substance/disease relationship. A large number of evidence-based rating systems are currently in use today by physicians, dietitians and other health professionals. FDA has tentatively chosen to model its evidence-based rating system on that of the Institute for Clinical Systems Improvement (ICSI) as adapted by the American Dietetic Association with modifications specific to FDA. In making this tentative decision, FDA relied on criteria for evaluating evidence-based rating systems as reviewed and critiqued by the Agency for Healthcare Research and Quality (AHRQ). FDA also found the modifications from the American Dietetic Association to be particularly useful as they considered diet and health relationships, whereas other groups focused on drug and treatment applications.

---

13 Examples of evidence-based rating systems are described and evaluated in: Agency for Healthcare Research and Quality. Systems to rate the strength of scientific evidence. Evidence Report/Technology Assessment. Number 47, 2002. The Healthcare Research And Quality Act of 1999, Part B, Title IX, Section 911(a) mandated that the Agency for Healthcare Research and Quality (AHRQ), in collaboration with experts from the public and private sectors, identify methods or systems to assess health care research results, particularly "methods or systems to rate the strength of the scientific evidence underlying health care practice, recommendations in the research literature, and technology assessments."


15 The ICSI system has been adapted by the American Dietetic Association (ADA) for their evidence-based dietetics practice and, thus, the ADA modifications have addressed many of the diet/disease relationships that are also of interest to FDA. See: Myers EF, Pritchett E, Johnson EQ. Evidence-based practice guides vs. protocols: what's the difference? JADA. 2001;101:1085-1090.
How are “Rate” and “Rank” Used in this System?

The terms “rate” and “rank” are not used interchangeably to describe this system. The evaluation process involves three separate rating systems: (1) a rating for study design; (2) a rating for study quality; and (3) a rating for the strength of the entire body of evidence. Considering all classifications from the three rating systems, a final rank of the scientific evidence in support of a health claim would be assigned.

What are the Parts of an Evidence-based Rating System?

In order to evaluate the level of scientific support for a proposed substance/disease relationship, the Task Force recommends that the agency follow a six-part procedure.

Each part of the evidence-based rating system is described below:

1) Define the substance/disease relationship
   A proposed relationship between a substance and a disease or health-related condition is identified. If relevant, the subgroups within the general population, for which the relationship is targeted are identified. The relationship forms the basis for selecting relevant studies and for evaluating the quality of the selected studies.

2) Collect and submit all relevant studies
   All relevant studies (both favorable and unfavorable) to the relationship to be tested (as defined above in 1) are collected and submitted. The evaluation of the proposed relationship relies primarily on human studies.

3) Classify, and therefore rate, each study as to type of study
   Each study would be characterized as a study design type. By categorizing the study, it automatically receives an initial study “rating” based on the type of experimental design, which is independent of the quality of the study. The rating of study design is based primarily on the principle of minimizing bias. Only primary reports of data collection are rated. Reports that synthesize or reflect...
collections of primary reports are not considered part of the rating system although they may provide useful background information.

(a) Study Design Type One
   - Randomized, controlled intervention trials

(b) Study Design Type Two
   - Prospective observational cohort studies

(c) Study Design Type Three
   - Nonrandomized intervention trials with concurrent or historical controls
   - Case-control studies

(d) Study Design Type Four
   - Cross-sectional studies
   - Analyses of secondary disease endpoints in intervention trials
   - Case series

4) Rate each study for quality

Each study would be reviewed independently and assigned a quality factor of +, $\emptyset$, -, or N/A. The basis for the assignment of the quality factor is discussed below.\textsuperscript{19}

(a) (+) means the report has adequately addressed issues of scientific quality such as inclusion/exclusion, bias, generalizability, and data collection and analysis.

(b) ($\emptyset$) means some uncertainties exist as to whether the report has adequately addressed issues of scientific quality such as inclusion/exclusion, bias, generalizability, and data collection and analysis.

(c) (-) means the report has not adequately addressed issues of scientific quality such as inclusion/exclusion, bias, generalizability, and data collection and analysis.

(d) N/A means the report is not a primary reference, therefore the quality has not been assessed, and such a reference is not considered as part of the body of evidence on which the final ranking is based. Examples of non-primary references are review articles and meta analyses.

5) Rate the strength of the total body of evidence.

The studies are considered collectively across the evidence base in order to rate the strength of the body of evidence. The rating system would be based on three factors: quantity, consistency, and relevance to disease risk reduction in the general population or target subgroup. These three factors and the final "rank" for the strength of the evidence for the "relationship" are described below.

\textsuperscript{19} Additional specific, detailed criteria, based on the above noted general principles, will be evaluated for usefulness during this interim period.
(a) Rating the body of evidence for quantity, consistency, and relevance to
disease risk reduction in the general population or target subgroup.

1. **Quantity.** Considers the number of studies, the total number of
individuals studied and the generalizability of the findings to the target
population.
   a. (***) means the number of studies and the number of individuals
tested (from all studies of design types one and two that are of
high quality (+) combined) are sufficiently large to comfortably
generalize to the target population.
   b. (**) means there are a sufficient number of studies and
individuals tested from study design type three and higher (i.e.
study design types one and two) of at least moderate quality (0)
but uncertainties remain as to generalizability to the target
population.
   c. (*) means that the number of studies and the number of
individuals tested is insufficient to generalize to the target
population.

2. **Consistency.** Considers whether studies with both similar and different
designs report similar findings.
   a. (***) means a sufficient number of studies of design types one
and two that are of high quality (+) have consistent results. Any
inconsistencies should be explained satisfactorily.
   b. (**) means there is a moderate consistency across all study
levels.
   c. (*) means that the results of studies are inconsistent.

3. **Relevance to Disease Risk Reduction in the General Population or Target
Subgroup.** Considers whether or not the magnitude of the risk-reduction
effect in the target population is physiologically meaningful and achievable in
the general US population or a subgroup of the US general population under
intake and use conditions that are appropriate for such conventional human
food and human dietary supplements that would be the subject of the claim.
   a. (***) means that the magnitude of the effect observed in studies
of design types one and two that are of high quality (+) is
physiologically meaningful and achievable under intake and use
conditions that are appropriate for such conventional human
food and human dietary supplements that would be the subject
of the claim.
   b. (**) means there is some suggestion from studies of design type
three and higher (i.e. study design types one and two) and of
moderate (0) to high (+) quality that the effect will be
physiologically meaningful, and achievable under intake and
use conditions that are appropriate for such conventional human food and human dietary supplements that would be the subject of the claim but uncertainties remain.

(*) means that the magnitude of the effect in the studies is not likely to be physiologically meaningful or achievable under intake and use conditions that are appropriate for such conventional human food and human dietary supplements that would be the subject of the claim.
(b) Ranking the Strength of the Evidence for a Health Claim

1. The first level, or highest rank of scientific evidence to support the substance/disease relationship, meets the "Significant Scientific Agreement among qualified experts" standard. (For the purpose of this guidance, the first level rank is used only as a reference point. In all other respects it is outside the scope of this guidance.)

This level reflects a high level of comfort among qualified scientists that the claimed substance/disease relationship is scientifically valid. In general, the first level ranked relationship would be considered to have a very low probability of significant new data overturning the conclusion that the relationship is valid or significantly changing the nature of the relationship. It would have high consistency with conclusions of authoritative bodies. The relationship would be based on relevant, high quality studies of mostly study design types one and two, and sufficient numbers of individuals would be tested to result in a high degree of confidence that results are relevant to the target population. Studies of different design would almost always result in similar findings, and the benefit would be physiologically meaningful and achievable under intake and use conditions that are appropriate for such conventional human food and human dietary supplements that would be the subject of the claim.

2. The second level rank of scientific evidence to support the substance/disease relationship is the highest level for a qualified health claim, and represents a moderate/good level of comfort among qualified scientists that the claimed relationship is scientifically valid. Qualified experts would rank the relationship as "promising," but not definitive. The claim would be based on relevant, high to moderate quality studies of study design type three and higher (i.e. design types one and two) and sufficient numbers of individuals would be tested to result in a moderate degree of confidence that results could be extrapolated to the target population. Studies of similar or different design would generally result in similar findings and the benefit would reasonably be considered to be physiologically meaningful and achievable under intake and use conditions that are appropriate for such conventional human food and dietary supplements that would be the subject of the claim. (Note: The term "moderate/good" for the second level rank may seem ungenerous. This terminology derives from historical data evaluated by the National Academy of Sciences that indicated that over time many

---

20 The use of the phrase "level of comfort" is mentioned in rulemaking that established the general requirements for health claims (21 CFR 101.14), which published in the Federal Register (58 FR 2478 at 2506; January 6, 1993)

diet/disease relationships that met this level of evidence were not necessarily sustained.)

3. The third level rank of scientific evidence to support the substance/disease relationship is the middle level for a qualified health claim and represents a low level of comfort among qualified scientists that the claimed relationship is scientifically valid. It would have low consistency with statements from authoritative bodies or be ranked as "low" in terms of scientific support by qualified scientists. The relationship would be based mostly on moderate to low quality studies of study design type three, and insufficient numbers of individuals would be tested, resulting in a low degree of confidence that results could be extrapolated to the target population. Studies of different design would generally result in similar findings but uncertainties would exist. Uncertainties would also exist as to whether the benefit would be considered physiologically meaningful and achievable under intake and use conditions that are appropriate for such conventional human food and human dietary supplements that would be the subject of the claim.

4. The fourth level, or the lowest rank of scientific evidence to support the claimed substance/disease relationship, is the lowest level for a qualified health claim and represents an extremely low level of comfort among qualified scientists that the claimed relationship is scientifically valid. It would have very low consistency with conclusions of authoritative bodies or be ranked very low by qualified scientists. The relationship would be based mostly on moderate to low quality studies of study design type three and insufficient numbers of individuals would be tested, resulting in a very low degree of confidence that results could be extrapolated to the target population. Studies of different design would generally result in similar findings but uncertainties would exist. There could be considerable uncertainty as to whether or not the benefit would be considered physiologically meaningful or achievable under intake and use conditions that are appropriate for such conventional human food and human dietary supplements that would be the subject of the claim. This level requires at least some credible evidence to support the relationship. There cannot be a strong body of evidence against the claim (e.g., a study or studies of high persuasiveness, quality and relevance that do not detect an effect). If that is the case, such evidence provides a sound basis for concluding that the claim is not valid.

5. If the scientific evidence to support the substance/disease relationship is below that described as the fourth level (see above) no claim would be appropriate.

6) Report the rank

The result of the evidence-based rating system would be a statement describing the nature of the evidence and the rationale for linking a substance to a
Attachment B – Interim Evidence-based Ranking System for Scientific Data

disease/health-related condition with a ranking as to the strength of the scientific evidence in support of that relationship. The process for arriving at the rank of the evidence to support the substance/disease relationship is illustrated in Table 1. The rank would be supported by:

(a) A clear and transparent demonstration of which research studies were evaluated to provide the rank.

(b) Evidence tables showing the rigor of the evaluation.
Attachment B – Interim Evidence-based Ranking System for Scientific Data

Table 1. Overview of the evidence-based rating system for evaluating the substance/disease relationship that is the subject of a qualified health claim. There are six steps to evaluating the strength of the scientific evidence in support of a qualified health claim.

<table>
<thead>
<tr>
<th>Step One</th>
<th>A proposed relationship between a substance and a disease or health-related condition is identified.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step Two</td>
<td>Individual studies are identified that are pertinent to the substance/disease relationship.</td>
</tr>
<tr>
<td>Step Three</td>
<td>Individual studies are classified according to study design type. Different design types are graded higher than others, based on their ability to minimize bias. Thus assignment of a study design automatically provides a rating.</td>
</tr>
<tr>
<td>Step Four</td>
<td>Individual studies are assigned a designation of +, 0, -, or N/A to reflect the study quality. (The general criteria for quality determination are described in this attachment).</td>
</tr>
<tr>
<td>Step Five</td>
<td>The strength of the scientific evidence in support of the substance/disease relationship is given a rank. This rank is determined taking into account the quantity, consistency, and relevance to disease risk reduction of the aggregate of the studies.</td>
</tr>
<tr>
<td>Step Six</td>
<td>The rank is reported.</td>
</tr>
</tbody>
</table>
What Resource Materials are Available?

1. Internet-based Resource Materials

- Agency for Healthcare Research and Quality (at http://www.ahrq.gov)
- American Dietetic Association (at http://www.eatright.org/)
- Canadian Task Force on Preventive Health Care (at http://www.ctfphc.org/)
- Center for Evidence Based Medicine (at http://www.cebm.utoronto.ca)
- Cochrane Collaboration/Cochrane Reviews (at http://www.cochrane.org)
- Evidence-based Practice Internet Resources (at http://www-hsl.mcmaster.ca/ebm)
- Federal Judicial Center (at http://www.fjc.gov)
- Health Canada. Since their June 2000 publication of the proposed standards for health claims, proposals on two approaches to regulating health claims on foods have been published. The two approaches are: generic authorization and product-specific authorization (see http://canada.ca).
- National Coordination Centre for Health Technology Assessment (at http://www.ncchta.org/main.htm)
- National Health and Medical Research Council (at http://www.health.gov.au/nhmrc/)
- National Health Service Centre for Reviews and Dissemination (http://www.york.ac.uk/inst/crd/).
Attachment B – Interim Evidence-based Ranking System for Scientific Data

- New Zealand Guidelines Group (at http://www.nzgg.org.nz/)

- Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't (see http://www.cebm.net/ebm_is_isnt.asp)

- Scottish Intercollegiate Guidelines Network (at http://www.sign.ac.uk/)

2. Other Resource Materials


Attachment B – Interim Evidence-based Ranking System for Scientific Data


- Guyatt GH, Sackett DL, Cook DJ. Users’ guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-based medicine working group. *JAMA.* 1993; 270:2598-2601.


Attachment B – Interim Evidence-based Ranking System for Scientific Data


