

Perspective: USDA Nutrition Evidence Systematic Review Methodology: Grading the Strength of Evidence in Nutrition- and Public Health–Related Systematic Reviews

Maureen K Spill,^{1,2} Laural K English,^{1,2} Ramkripa Raghavan,^{1,2} Emily Callahan,² Darcy Güngör,^{1,2} Brittany Kingshipp,^{1,2} Joanne Spahn,² Eve Stoodly,³ and Julie Obbagy²

¹Panum Group, Bethesda, MD, USA; ²Nutrition Evidence Systematic Review (NESR) team, Nutrition Guidance and Analysis Division (NGAD), Center for Nutrition Policy and Promotion (CNPP), Food and Nutrition Service (FNS), USDA, Alexandria, VA, USA; and ³Nutrition Guidance and Analysis Division (NGAD), Center for Nutrition Policy and Promotion (CNPP), Food and Nutrition Service (FNS), USDA, Alexandria, VA, USA

ABSTRACT

The USDA's Nutrition Evidence Systematic Review (NESR) team conducts food- and nutrition-related systematic reviews used to inform US Federal guidelines and programs, including the *Dietary Guidelines for Americans*. NESR's systematic review methodology includes a step to grade the strength of the evidence underlying conclusion statements, which is critical for ensuring that end users understand the level of certainty in conclusions when using them to make decisions. Over time, NESR has ensured its grading process not only remains state of the art but is also designed specifically for systematic reviews that inform Federal guidelines and programs on nutrition and public health. The NESR grading process used by the 2020 Dietary Guidelines Advisory Committee included 5 grading elements: risk of bias, consistency, directness, precision, and generalizability. Evidence was grouped by study design and assessed against these elements, and the grade assigned to the entire body of evidence took into consideration the strengths and limitations of each design. Based on this assessment, 1 of 4 grades was assigned: strong, moderate, limited, or grade not assignable. The grade was clearly communicated by integrating specific language into each conclusion statement (e.g., "strong evidence demonstrates" or "limited evidence suggests"), and supported by rationale documented in the review. NESR's grading process aligns with approaches used by other organizations that conduct systematic reviews, while retaining aspects unique to NESR's role in informing Federal nutrition and public health guidelines and programs. It provides a framework that promotes consistency in grading across food- and nutrition-related reviews, while offering flexibility that allows for thorough consideration of the body of evidence underlying an individual conclusion statement. NESR's rigorous and transparent methods for grading the strength of evidence in food- and nutrition-related systematic reviews ensure that decisions related to nutrition and public health are based on the strongest available evidence. *Adv Nutr* 2022;13:982–991.

Statement of Significance: This article describes the rigorous and transparent methods developed and used by USDA's Nutrition Evidence Systematic Review (NESR) team to grade the strength of evidence in food- and nutrition-related systematic reviews used to inform US Federal nutrition guidelines and programs. NESR's grading process was used in systematic reviews conducted by the 2020 Dietary Guidelines Advisory Committee, which informed the development of the *Dietary Guidelines for Americans, 2020–2025*.

Keywords: grading, NESR, nutrition evidence systematic review, strength of evidence, certainty of evidence, systematic review methodology, systematic reviews, Dietary Guidelines for Americans, public health, federal food policy and programs

Introduction

The USDA's Nutrition Evidence Systematic Review (NESR) team specializes in conducting food- and nutrition-related systematic reviews. NESR systematic reviews are research projects that answer important public health questions by using rigorous and transparent methods to search for,

evaluate, analyze, and synthesize the body of scientific evidence on topics relevant to US Federal policy and programs. NESR is housed within the USDA's Center for Nutrition Policy and Promotion (CNPP), and its work supports CNPP's mission to improve the health of Americans by developing and promoting dietary guidance that links scientific research

to the nutrition needs of consumers. In addition, NESR's systematic review methodology promotes scientific integrity. Because of its evidence-based approach, NESR's work helps uphold the Data Quality Act (1), which mandates that Federal agencies ensure the quality, objectivity, utility, and integrity of the information used to form Federal guidance.

The NESR team collaborates with subject matter expert groups, such as the Dietary Guidelines Advisory Committee (the Committee), to conduct its systematic reviews using rigorous, protocol-driven methodology (2, 3). To ensure that NESR's process remains state of the art, NESR has routinely evaluated and refined its methodology through a continuous quality advancement (CQA) initiative. When appropriate, refinements to the NESR methodology are carefully adopted. Process refinements reflect continuous evolution in both systematic review and nutrition science. NESR's CQA initiative promotes efficiency and maintains the credibility and high quality of NESR's work.

NESR has always included grading the strength of evidence as part of its systematic review methodology, and through its CQA initiative has made changes over time to the processes and criteria used in grading, most recently in 2018 before supporting the 2020 Committee (3–5). These changes reflect evolutions in the field of systematic review methodology while providing consistency in NESR's grading process across time and projects (Table 1). As part of its 2018 CQA initiative, the NESR team reviewed and compared methods from several organizations and engaged with methodologists from some of those organizations (6–17). The NESR team also considered the National Academies of Sciences, Engineering, and Medicine (NASEM) report, "Redesigning the Process for Establishing the Dietary Guidelines for Americans" (18). This report provided NESR the following recommendation related to grading the strength of evidence:

"Conduct of original systematic reviews will need to be transparent and follow state-of-the-art methods, such as the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach and the AHRQ (Agency for Healthcare Research and Quality) Evidence-based Practice Centers Program approach. However, this National Academies committee believes the [NESR] and DGSAC (Dietary Guidelines Scientific

Supported by the USDA, Food and Nutrition Service, Center for Nutrition Policy and Promotion. Author disclosures: The authors report no conflicts of interest.

Perspective articles allow authors to take a position on a topic of current major importance or controversy in the field of nutrition. As such, these articles could include statements based on author opinions or point of view. Opinions expressed in Perspective articles are those of the author and are not attributable to the funder(s) or the sponsor(s) or the publisher, Editor, or Editorial Board of Advances in Nutrition. Individuals with different positions on the topic of a Perspective are invited to submit their comments in the form of a Perspectives article or in a Letter to the Editor.

Scientists who are employees of the funding source had a role in the design and conduct of this study; collection, management, analysis, and interpretation; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. Panum Group employees worked under contract with the Food and Nutrition Service, USDA. Address correspondence to JO (e-mail: julie.obbagy@usda.gov).

Abbreviations used: AHRQ, Agency for Healthcare Research and Quality; CNPP, Center for Nutrition Policy and Promotion; CQA, continuous quality advancement; GRADE, Grading of Recommendations Assessment, Development and Evaluation; NASEM, National Academies of Sciences, Engineering, and Medicine; NESR, Nutrition Evidence Systematic Review; RCT, randomized controlled trial.

TABLE 1 Characteristics of the Nutrition Evidence Systematic Review grading process across projects

	2010 Dietary Guidelines Advisory Committee Project (5)	Dietary Patterns Project (29)	2015 Dietary Guidelines Advisory Committee Project (4)	Pregnancy and Birth to 24 Months Project (3)	2020 Dietary Guidelines Advisory Committee Project (2)
Grading terminology	Strong Moderate Limited Expert opinion only Grade not assignable	Strong Moderate Limited Grade not assignable	Strong Moderate Limited Grade not assignable	Strong Moderate Limited Grade not assignable	Strong Moderate Limited Grade not assignable
Grading elements	Quality Quantity Consistency Magnitude of effect Generalizability	Quality Quantity Consistency Impact Generalizability	Risk of bias Quantity Consistency Impact Generalizability	Internal validity Adequacy Consistency Impact Generalizability	Risk of bias Directness Consistency Precision Generalizability Study design Publication bias
Other considerations					

Advisory Committee) need to have the flexibility to align with appropriate standards or methods and does not recommend any one standard be adopted, which may be subject to change and evolve over time."

The purpose of this article is to describe the NESR grading process, informed by the CQA initiative and NASEM recommendations, that was implemented in the systematic reviews conducted by the Committee (2). This Committee's review of science informed the *Dietary Guidelines for Americans, 2020–2025* (2), which provides food-based recommendations to promote health, help prevent diet-related disease, and meet nutrient needs.

The NESR Grading Process Used for Systematic Reviews Conducted by the 2020 Committee

NESR's systematic review methodology

The NESR team used a rigorous, protocol-driven methodology to support the Committee in conducting systematic reviews, and a detailed description of the methodology can be found in the scientific report (2). In short, NESR's methodology for answering a systematic review question involves:

- developing a protocol,
- searching for and screening all relevant articles of primary research,
- extracting data and assessing the risk of bias of results from each included article,
- synthesizing the evidence,
- developing conclusion statements,
- grading the evidence underlying each conclusion statement, and
- recommending future research.

The grade is a critical part of the process because it communicates the strength of the evidence underlying a specific conclusion statement to decision makers and stakeholders. Specifically, in the case of reviews conducted by the Committee, graded conclusion statements were an important resource used in developing the *Dietary Guidelines for Americans, 2020–2025* (2).

NESR's grading process

The process of grading the strength of evidence underlying a conclusion statement begins by examining the body of evidence by study design [e.g., randomized controlled trials (RCTs), prospective cohort studies]. This allows for careful consideration of the strengths and limitations of each study design included in the review. Study design is further taken into account through the grading element of risk of bias, described below, because NESR's process utilizes risk-of-bias tools that capture design-specific concerns.

The evidence from each respective design is assessed based on 5 grading elements: risk of bias, consistency, directness, precision, and generalizability. This assessment is facilitated using the NESR grading rubric (Table 2), which is publicly available on the NESR website

(<https://nesr.usda.gov/2020-dietary-guidelines-advisory-committee-systematic-reviews>) (2). The grading elements represent related but separate concepts:

- 1) **Risk of bias** considers the likelihood that systematic errors resulting from the design and conduct of the studies could have affected the accuracy of the reported results across the body of evidence. In NESR's process, each included study undergoes a formal risk-of-bias assessment specific to the study design (2). These assessments of each individual study are documented, compiled, and used to inform consideration of risk of bias across the body of evidence.
- 2) **Consistency** considers the degree of similarity in the direction and magnitude of effect across the body of evidence. This element also considers whether differences across the results can be explained by variations in study designs and methods. For example, the differences in the magnitude of effects across studies may be attributed to differences in population (e.g., baseline nutritional status), the intervention/exposure (e.g., dosage), or other study methods.
- 3) **Directness** considers the extent to which studies are designed to directly examine the relation among the interventions/exposures, comparators, and outcome of primary interest in the systematic review question.
- 4) **Precision** considers the degree of certainty around an effect estimate for a given outcome. This element considers measures of variability, such as the width and range of CIs, the number of studies, and sample sizes, within and across studies.
- 5) **Generalizability** considers whether the study participants, interventions and/or exposures, comparators, and outcomes examined in the body of evidence are applicable to the US population of interest for the review.

The assessments of each grading element, made by study design using the NESR grading rubric, are then used to facilitate the expert group's discussion and selection of an overall grade for the entire body of evidence, across all included study designs, underlying the conclusion statement. The overall grade is not based on a predefined formula for scoring or tallying ratings of each element. Rather, each overall grade reflects the experts' thorough consideration of all of the elements, as they each relate to the specific nuances of the body of evidence under review. In the text of the report, the experts describe their assessment of the grading elements which provides a transparent explanation of the overall grade selected. This article provides examples from the 2020 Committee to illustrate this point.

Grades of strong, moderate, or limited may be given to a NESR conclusion statement. If a conclusion statement is not able to be drawn owing to there being no or insufficient evidence, no grade is assigned (i.e., grade not assignable). Table 3 provides definitions of the grades used in the NESR process. The grade is clearly communicated in a standardized manner across all conclusion statements, such

TABLE 2 Nutrition Evidence Systematic Review grading rubric used by the 2020 Dietary Guidelines Advisory Committee

Grading elements	Strong	Moderate	Limited	Grade not assignable
Risk of bias	Across the body of evidence, there is a <u>strong</u> likelihood that the design and conduct of the studies has prevented or minimized bias such that the reported results are the true effects of the intervention/exposure, and plausible bias and/or potential limitations are unlikely to alter the results	Across the body of evidence, there is a <u>moderate</u> likelihood that the design and conduct of the studies has prevented or minimized bias such that the reported results are the true effects of the intervention/exposure, and plausible bias and/or potential limitations are unlikely to alter the results	Across the body of evidence, there is a <u>limited</u> likelihood that the design and conduct of the studies has prevented or minimized bias such that the reported results may not be the true effects of the intervention/exposure, and plausible bias and/or potential limitations may have altered the results	A grade is not assignable for this element because it cannot be adequately assessed
Consistency	The body of evidence demonstrates findings with <u>strong</u> consistency in direction and magnitude of effect; or, any inconsistencies in findings can be explained by methodological differences	The body of evidence demonstrates findings with <u>moderate</u> consistency in direction and magnitude of effect; some of the inconsistencies in findings can be explained by methodological differences	The body of evidence demonstrates findings with <u>limited</u> consistency in direction and magnitude of effect; few of the inconsistencies in findings can be explained by methodological differences	A grade is not assignable for this element because it cannot be adequately assessed
Directness	The body of evidence demonstrates <u>strong</u> directness, such that studies are designed to directly examine the relations among intervention/exposure, comparator, and outcomes of primary interest in the systematic review question	The body of evidence demonstrates <u>moderate</u> directness, such that some studies are designed to directly examine the relations among intervention/exposure, comparator, and/or outcomes of primary interest in the systematic review question	The body of evidence demonstrates <u>limited</u> directness, such that few studies are designed to directly examine the relations among intervention/exposure, comparator, and/or outcomes of primary interest in the systematic review question	A grade is not assignable for this element because it cannot be adequately assessed
Precision	The body of evidence demonstrates <u>strong</u> precision based on a substantial number of sufficiently powered studies with a narrow assessment of variance	The body of evidence demonstrates <u>moderate</u> precision based on an adequate number of sufficiently powered studies with a narrow assessment of variance	The body of evidence demonstrates <u>limited</u> precision based on an inadequate number of sufficiently powered studies with a narrow assessment of variance	A grade is not assignable for this element because it cannot be adequately assessed
Generalizability	The body of evidence demonstrates <u>strong</u> generalizability to the US population of interest with regard to: a) the participant characteristics; b) the intervention/exposure and outcomes studied	The body of evidence demonstrates <u>moderate</u> generalizability to the US population of interest with regard to: a) the participant characteristics; b) the intervention/exposure and outcomes studied	The body of evidence demonstrates <u>limited</u> generalizability to the US population of interest with regard to: a) the participant characteristics; b) the intervention/exposure and outcomes studied	A grade is not assignable for this element because it cannot be adequately assessed

TABLE 3 Nutrition Evidence Systematic Review grades and definitions used by the 2020 Dietary Guidelines Advisory Committee

<u>Strong</u>	The conclusion statement is based on a strong body of evidence as assessed by risk of bias, consistency, directness, precision, and generalizability. The level of certainty in the conclusion is strong, such that if new evidence emerges, modifications to the conclusion are unlikely to be required
<u>Moderate</u>	The conclusion statement is based on a moderate body of evidence as assessed by risk of bias, consistency, directness, precision, and generalizability. The level of certainty in the conclusion is moderate, such that if new evidence emerges, modifications to the conclusion may be required
<u>Limited</u>	The conclusion statement is based on a limited body of evidence as assessed by risk of bias, consistency, directness, precision, and generalizability. The level of certainty in the conclusion is limited, such that if new evidence emerges, modifications to the conclusion are likely to be required
<u>Grade not assignable</u>	A conclusion statement cannot be drawn due to either a lack of evidence, or evidence that has severe limitations related to risk of bias, consistency, directness, precision, and/or generalizability

that the beginning of the statement identifies the strength of the evidence by the grade. The verb that follows is also tied to the grade, and is intended to communicate certainty in the conclusion. Conclusion statements use the language “strong evidence demonstrates,” “moderate evidence indicates,” or “limited evidence suggests.”

The grading process is often iterative because it relies on careful consideration of the body of evidence and the drafted conclusion statement. During the grading process, a conclusion statement may be modified to more accurately represent the evidence, such as including a specific population group or outcome measure. In turn, the grade is re-evaluated to ensure it reflects the updated conclusion statement.

Implementing the NESR grading process for the 2020 Committee

Training and guidance on the NESR grading process were provided to the Committee before commencing the systematic reviews. First, the Committee was oriented to NESR’s methodology during Committee meetings, which were open to the public, as well as administrative training meetings. In addition, NESR analysts provided training and guidance to the Committee throughout the duration of their work to ensure that grading was executed consistently in accordance with the NESR methodology.

During the Committee’s process to grade the evidence, NESR analysts provided the information necessary for assessing each grading element, including extracted data, risk of bias assessments, a description of the body of evidence, and a drafted synthesis of the evidence. NESR analysts also provided support during the Committee’s discussions on the body of evidence to ensure that all grading elements were discussed, that the conclusions and grades reflected the evidence reviewed, and that the rationale was transparently documented.

During public meetings, members of the Committee who supported a specific NESR systematic review presented and discussed the full systematic review, including the graded conclusion statements. For each NESR systematic review, the Committee described the evidence, explained the rationale for the grade, and addressed comments and questions from fellow Committee members. The public could participate in

the process by submitting comments, which were reviewed and considered by the Committee. Finally, all systematic reviews were peer reviewed by government scientists (2). These layers of input further strengthened the grading process.

Examples from the 2020 Committee

The NESR grading process provides a framework that allows for the strength of the evidence to be rated in a systematic and transparent manner across NESR systematic reviews. This framework does not offer a “one size fits all” approach, but rather, it promotes consistency in the approach to grading, while offering flexibility that allows each body of evidence to be weighed against all of the grading elements. Despite the unique nature of the body of evidence underlying a conclusion statement, each grade conveys the same meaning regardless of the rationale (Table 3). Furthermore, systematic review questions can be answered with more than 1 conclusion statement, each assigned its own grade. Therefore, grades of conclusion statements from the same systematic review question may differ. To illustrate these important concepts, examples of graded conclusion statements from systematic reviews conducted by the Committee are provided below (Tables 4 and 5).

Conclusion statements with the same grade may have different rationales.

Conclusion statements that received the same grade may differ in terms of the underlying bodies of evidence, such as in number of studies or study designs. The rationales for the same grade may also differ depending on the assessment of the evidence against each of the grading elements. However, the meaning of the grade is the same (Table 3) regardless of these differences. For example, each of the following conclusion statements was graded moderate (Table 4):

- Moderate evidence indicates dietary patterns higher in vegetables, fruits, legumes, whole grains, lean meats and seafood, and low-fat dairy; and low in red and processed meats, saturated fat, and sodas and sweets relative to other dietary patterns are associated with lower risk of colon and rectal cancer. Moderate

TABLE 4 Examples of conclusion statements with the same grade¹

Systematic review questions	What is the relationship between dietary patterns consumed and certain types of cancers?	What is the relationship between folic acid from supplements and/or fortified foods consumed before and during pregnancy and lactation and health outcomes?
Grade	Moderate	Moderate
Conclusion statement	Moderate evidence indicates dietary patterns higher in vegetables, fruits, legumes, whole grains, lean meats and seafood, and low-fat dairy; and low in red and processed meats, saturated fat, and sodas and sweets relative to other dietary patterns are associated with lower risk of colon and rectal cancer. Moderate evidence also indicates dietary patterns that are higher in red and processed meats, French fries and potatoes, and sources of sugars (i.e., sodas, sweets, and dessert foods) are associated with a greater colon and rectal cancer risk	Moderate evidence indicates that folic acid supplements consumed during lactation are positively associated with red blood cell folate, and may be positively associated with serum or plasma folate
Body of evidence	24 articles: 2 RCTs, 21 prospective cohort studies, 1 nested case-control	5 articles: 3 RCTs, 1 prospective cohort study, 1 uncontrolled before-and-after
Risk of bias	Some concerns related to potential confounding and possible changes in dietary intake over follow-up	Some concerns related to potential bias in selection of results in RCTs, and serious concerns related to confounding
Consistency	Consistency was demonstrated with the majority of studies reporting statistically significant relations between a "healthy" dietary pattern and lower all-cause mortality risk	Consistency demonstrated with supplementation associated with higher values on ≥ 1 measure of red blood cell folate in each study
Directness	Few or no concerns	Few or no concerns
Precision	Few or no concerns	Few or no concerns
Generalizability	Few or no concerns	Some concerns owing to a lack of diversity in participant characteristics; most participants were from relatively high socioeconomic backgrounds, with little racial/ethnic diversity reported

¹RCT, randomized controlled trial.

evidence also indicates dietary patterns that are higher in red and processed meats, French fries and potatoes, and sources of sugars (i.e., sodas, sweets, and dessert foods) are associated with a greater colon and rectal cancer risk. (Grade: Moderate) (19).

- Moderate evidence indicates that folic acid supplements consumed during lactation are positively associated with red blood cell folate, and may be positively associated with serum or plasma folate. (Grade: Moderate) (20).

Although both of these conclusion statements were graded moderate and the evidence underlying each had few or no concerns related to directness and precision, there were several differences worth noting. One of the differences is the number and types of studies underlying each conclusion (Table 4). The first example on dietary patterns was informed by 24 articles, including 2 RCTs, 21 prospective cohort studies, and 1 nested case-control study. The second example on folic acid supplementation was informed by 5 articles from 3 RCTs, 1 prospective cohort study, and 1 uncontrolled before-and-after study. In addition, there were differences in risk of bias, consistency, and generalizability. The assessment of risk of bias for the dietary patterns evidence identified concerns related to potential confounding and possible changes in dietary intake over follow-up, whereas the assessment of the folic acid evidence identified concerns

related to selection of results in the RCTs and potential confounding in the other studies. For the dietary patterns evidence, there was good consistency in that the majority of studies reported statistically significant relations between dietary patterns consumed and colon and rectal cancer risk. For folic acid, results were highly consistent, with each of the studies reporting that supplementation was associated with higher values on ≥ 1 measure of red blood cell folate. The assessment of generalizability for the dietary patterns evidence identified few or no concerns, whereas the assessment of the folic acid evidence identified some concerns, given that participants were from relatively high socioeconomic backgrounds, with little racial or ethnic diversity reported. Although each body of evidence had different concerns, both reflected a moderate level of certainty, such that if new evidence emerges, modifications to those conclusions may be required.

Conclusion statements from the same question may have different grades.

A single systematic review question can be answered with multiple graded conclusion statements, each of which addresses different parts of the question (e.g., population subgroups, intervention details, outcome measures). And, different conclusion statements developed as part of the same

TABLE 5 Examples of conclusion statements with different grades¹

Systematic review question: What is the relationship between folic acid from supplements and/or fortified foods consumed before and during pregnancy and risk of hypertensive disorders during pregnancy?			
Intervention/exposure	Folic acid supplements	Folic acid supplements	Folic acid from fortified foods
Population	Women at high risk	Women at low risk	All women
Grade	Limited	Moderate	Grade not assignable
Conclusion statement	Limited evidence suggests that folic acid supplements consumed during early pregnancy may have a beneficial effect on reducing the risk of hypertensive disorders during pregnancy among women at high risk (e.g., history of pre-eclampsia or prepregnancy BMI ≥ 25 kg/m ²) compared with no folic acid supplementation	Moderate evidence indicates that higher levels of folic acid supplements consumed during pregnancy compared with lower levels (including no folic acid supplementation) does not affect the risk of hypertensive disorders during pregnancy among women at low risk	No evidence is available to determine the relation between folic acid from fortified foods consumed before and during pregnancy and the risk of hypertensive disorders during pregnancy
Body of evidence	5 articles: 2 nonrandomized controlled trials, 3 prospective cohort studies	6 articles: 3 randomized controlled trials, 3 prospective cohort studies	0 articles
Risk of bias	Concerns about design and conduct of nonrandomized controlled trials; classification of exposures deviation from intended exposure of prospective cohort studies	Concerns of randomization method for 1 RCT, missing data from another RCT; classification of exposures deviation from intended exposure of prospective cohort studies	Not applicable
Consistency	Few to no concerns	Few to no concerns	Not applicable
Directness	Few to no concerns	Few to no concerns	Not applicable
Precision	Concerns related to small sample sizes and number of cases, differences in prevalence across studies, and wide CIs	Concerns related to small sample sizes and number of cases, and concerns that studies were not sufficiently powered	Not applicable
Generalizability	Concerns related to little racial/ethnic diversity and little data provided on other participant characteristics	Concerns related to little racial/ethnic diversity and little data on other participant characteristics; RCTs only recruited healthy women from Iran	Not applicable

¹RCT, randomized controlled trial.

systematic review may receive different overall grades reflecting the evidence supporting those statements; the grades do not reflect evidence that is irrelevant to the conclusion statements. For example, a systematic review was conducted to answer the question, “What is the relationship between folic acid from supplements and/or fortified foods consumed before and during pregnancy and risk of hypertensive disorders during pregnancy?” Eight studies were included in the review: 3 RCTs, 2 nonrandomized controlled trials, and 3 prospective cohort studies (Table 5). After thoroughly reviewing the evidence, the systematic review question was answered with 3 conclusion statements, each of which was supported by a subset of those 8 studies. The conclusions were:

- Limited evidence suggests that folic acid supplements consumed during early pregnancy may have a beneficial effect on reducing the risk of hypertensive disorders during pregnancy among *women at high risk* (e.g., history of pre-eclampsia or prepregnancy BMI ≥ 25 kg/m²) compared with no folic acid supplementation. (Grade: Limited) (20).

- Moderate evidence indicates that higher levels of folic acid supplements consumed during pregnancy compared with lower levels (including no folic acid supplementation) does not affect the risk of hypertensive disorders during pregnancy among *women at low risk*. (Grade: Moderate) (20).
- No evidence is available to determine the relation between folic acid from fortified foods consumed before and during pregnancy and the risk of hypertensive disorders during pregnancy. (Grade: Grade not assignable) (20).

Evidence supporting the conclusion related to consumption of folic acid supplements by *women at high risk* was assigned a grade of limited. It included the 2 nonrandomized control trials and 3 prospective cohort studies, but not the 3 RCTs, which only recruited healthy women at low risk of hypertensive disorders. The evidence supporting this conclusion demonstrated good consistency and directly examined the exposure, comparator, and outcomes of interest. However, owing to a lack of evidence from RCTs and concerns related to risk of bias, precision, and

generalizability, there was limited certainty in the conclusion. Evidence supporting the conclusion related to consumption of folic acid supplements by women at low risk was assigned a grade of moderate. This body of evidence included the 3 RCTs and 3 prospective cohort studies. This evidence was similar to that in high-risk women, in that it also had good consistency and directness, with some concerns related to precision and generalizability. However, this evidence received a grade of moderate, rather than limited, mostly owing to the inclusion of consistent evidence from RCTs and fewer risk-of-bias concerns. Finally, no evidence was available that met the inclusion criteria and addressed the relation between folic acid from fortified foods consumed before and during pregnancy and the risk of hypertensive disorders during pregnancy, resulting in “grade not assignable.”

Extensive thought and consideration go into the determination of each grade, and no 2 sets of evidence are the same. Each evidence base is evaluated independently and specifically in relation to the conclusion statement it supports. The NESR grading process provides structure to assess every evidence base, regardless of differences, and consistency in terminology to ensure that grades are interpreted in a similar way.

Discussion

NESR's grading process aligns with those used by other organizations, such as the GRADE approach and that of AHRQ (9, 21). NESR and these others rely on consideration of specific elements (2, 9, 21), sharing 4 of 5 grading elements in common: risk of bias, consistency, directness, and precision (9, 22–26). All of these approaches take study design into consideration, and all assign an overall grade that communicates the strength of the evidence to decision makers and stakeholders (27). NESR's grading process differs in its consideration of study design, publication bias, and generalizability, given its role in informing public health nutrition decisions within the US government.

NESR systematic reviews include study designs that offer the strongest evidence with which to establish a relationship, including both RCT and nonrandomized designs. NESR recognizes that relying on RCTs is important, but also considers that rigorously conducted nonrandomized studies, such as prospective cohort studies, can provide important evidence that complements data from RCTs, particularly when RCTs are not feasible. The NESR grading process therefore considers the strengths and weaknesses of evidence within and across all included study designs. This process differs from others that begin with a set grade based on study design (e.g., RCT evidence begins with a grade of strong), and then adjust up or down based on the assessment of evidence (9, 21).

NESR's systematic review methodology includes steps to assess the potential for publication bias, but publication bias is not a formal element in NESR's grading process. NESR recognizes the importance of publication bias in nutrition research and carefully evaluates and documents it during evidence synthesis and assessment. Publication bias may not

effectively differentiate the strength of bodies of evidence, and there are not widely accepted, reliable methods for assessing publication bias and its impact on systematic review conclusions (9, 24, 28).

NESR's grading process has always included generalizability as a formal grading element. Because NESR systematic reviews are conducted to inform Federal nutrition policies and programs in the United States, it is imperative that they consider how generalizable the evidence is to Americans at the population level.

Finally, NESR's grading process includes steps designed to transparently communicate the strength of evidence underlying each conclusion statement, which includes the grade followed by a specific complementary verb (e.g., “strong evidence demonstrates,” “limited evidence suggests”). This integration of the grade in the conclusion statement allows end users to clearly understand the certainty of evidence, and the likelihood it may change if new evidence emerges. NESR's process also accommodates situations in which no or insufficient evidence is available to develop a conclusion statement by allowing for “grade not assignable” to be used.

Continued advancement of NESR's grading process

NESR's grading process has evolved over time to remain state of the art, leveraging advances in both the fields of systematic review methodology and nutrition science. As noted in the NASEM report (18), appropriate standards or methods evolve and improve dynamically over time in response to advances in science. As part of NESR's commitment to quality advancement, the Committee provided feedback on their experience using NESR's systematic review process. Strengths identified included the utility of the NESR grading rubric, which was described by the Committee as objective, concise, and easy to use to assess grading elements. In addition, the Committee recommended that NESR publish its grading process to benefit others conducting nutrition-related systematic reviews. As part of NESR's CQA process, NESR staff will consider feedback from the Committee, continue to monitor advancements in grading approaches, and, when appropriate, update the NESR grading process.

Conclusion

NESR's grading process provided a structured and transparent approach to assessing the strength of evidence underlying conclusion statements drawn as part of the systematic reviews conducted by the Committee on relations between nutrition and health outcomes. The process relied on consideration of 5 grading elements (risk of bias, consistency, directness, precision, and generalizability) as well as the study designs of included articles to determine and communicate the level of certainty in each conclusion drawn during the review process. The NESR grading process was designed to have specific utility in informing public health-related policy decisions relevant to Americans at the population level.

Over time, NESR's grading process has evolved, and it will continue to evolve to remain state of the art, leveraging

advances in both the fields of systematic review methodology and nutrition science. NESR's process aligns with methods used by other organizations that conduct systematic reviews to inform decision making, ensuring that decisions made related to nutrition and public health are based on the strongest available evidence.

Acknowledgments

We thank current and former staff on the Nutrition Evidence Systematic Review team including Charlotte Bahnfleth, Marlana Bates, Gisela Butera, Natasha Cole, Meredith Higgins, Julia Kim, Julie Nevins, Sara Scinto-Madonich, and Sudha Venkatramanan. The authors' responsibilities were as follows—MKS, LKE, RR, and JO: developed the NESR grading process described in this article and are responsible for the design, writing, and final content of the paper; EC, DG, BK, JS, and ES: provided feedback on the development of this grading process and provided feedback on the manuscript; and all authors: read and approved the final manuscript.

References

1. The Information Quality Act (IQA), Pub. L. No. 106-554, Section 515(a)(December 21, 2000).
2. Dietary Guidelines Advisory Committee. Scientific report of the 2020 Dietary Guidelines Advisory Committee: advisory report to the Secretary of Agriculture and the Secretary of Health and Human Services. Washington (DC): USDA Agricultural Research Service; 2020.
3. Obbagy JE, Spahn JM, Wong YP, Psota TL, Spill MK, Dreifelbis C, Gungor DE, Nadaud P, Raghavan R, Callahan EH, et al. Systematic review methods for the Pregnancy and Birth to 24 Months Project. *Am J Clin Nutr* 2019;109(Supplement_1):698S–704S.
4. Dietary Guidelines Advisory Committee. Scientific Report of the 2015 Dietary Guidelines Advisory Committee: Advisory Report to the Secretary of Health and Human Services and the Secretary of Agriculture [Internet]. Washington (DC): USDA Agricultural Research Service; 2015 [cited 24 June, 2020]. Available from: <https://health.gov/sites/default/files/2019-09/Scientific-Report-of-the-2015-Dietary-Guidelines-Advisory-Committee.pdf>.
5. Dietary Guidelines Advisory Committee. Report of the Dietary Guidelines Advisory Committee on the Dietary Guidelines for Americans, 2010, to the Secretary of Agriculture and the Secretary of Health and Human Services [Internet]. Washington (DC): USDA Agricultural Research Service; 2010 [cited 24 June, 2020]. Available from: <https://dietaryguidelines.gov/sites/default/files/2019-05/2010DGACReport-camera-ready-Jan11-11.pdf>.
6. Schünemann HJ. Interpreting GRADE's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? *J Clin Epidemiol* 2016;75:6–15.
7. Pena-Rosas JP, De-Regil LM, Rogers LM, Bopardikar A, Panisset U. Translating research into action: WHO evidence-informed guidelines for safe and effective micronutrient interventions. *J Nutr* 2012;142(1):197S–204S.
8. US Environmental Protection Agency (EPA). Exposure Factors Handbook 2011 edition (final report). Washington (DC): US EPA; 2011.
9. Berkman ND, Lohr KN, Ansari MT, Balk EM, Kane R, McDonagh M, Morton SC, Viswanathan M, Bass EB, Butler M, et al. Grading the strength of a body of evidence when assessing health care interventions: an EPC update. *J Clin Epidemiol* 2015;68(11):1312–24.
10. National Toxicology Program (NTP). Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration. Research Triangle Park, NC: NTP, National Institute of Environmental Health Sciences; 2015.
11. Woodruff TJ, Sutton P. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. *Health Aff* 2011;30(5):931–7.
12. World Cancer Research Fund/American Institute for Cancer Research. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington (DC): World Cancer Research Fund International; 2007.
13. Agency for Healthcare Research and Quality (AHRQ). U.S. Preventive Services Task Force procedure manual. Rockville, MD: AHRQ; 2008.
14. Centre for Reviews and Dissemination (CRD), University of York. Systematic reviews: CRD's guidance for undertaking reviews in health care. York, United Kingdom: CRD, University of York; 2009.
15. European Food Safety Authority (EFSA). Tools for critically appraising different study designs, systematic review and literature searches. *EFSA Support Publ* 2015;12(7):836E.
16. Academy of Nutrition and Dietetics. Evidence analysis manual: steps in the Academy evidence analysis process. Chicago, IL: Academy of Nutrition and Dietetics; 2016.
17. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, George SM, Olson RD. The Physical Activity Guidelines for Americans. *JAMA* 2018;320(19):2020–8.
18. National Academies of Sciences, Engineering and Medicine. Redesigning the process for establishing the Dietary Guidelines for Americans. Washington (DC): The National Academies Press; 2017.
19. Bouche Y, Ard J, Bazzano L, Heymsfield S, Mayer-Davis E, Sabaté J, Snetelaar L, Van Horn L, Schneeman B, English LK, et al. Dietary patterns and breast, colorectal, lung, and prostate cancer: a systematic review. Alexandria, VA: USDA, Food and Nutrition Service, Center for Nutrition Policy and Promotion, Nutrition Evidence Systematic Review; 2020.
20. Donovan S, Dewey K, Novotny R, Stang J, Taveras E, Kleinman R, Spill M, Kim JH, Nevins J, Raghavan R, et al. Folic acid from fortified foods and/or supplements during pregnancy and lactation and health outcomes: a systematic review. Alexandria, VA: USDA, Food and Nutrition Service, Center for Nutrition Policy and Promotion, Nutrition Evidence Systematic Review; 2020.
21. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64(4):383–94.
22. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol* 2011;64(12):1303–10.
23. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Glasziou P, Jaeschke R, Akl EA, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol* 2011;64(12):1294–302.
24. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, Alonso-Coello P, Djulbegovic B, Atkins D, Falck-Ytter Y, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64(12):1277–82.
25. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol* 2011;64(4):407–15.
26. Berkman ND, Lohr KN, Ansari M, McDonagh M, Balk E, Whitlock E, Reston J, Bass E, Butler M, Gartlehner G, et al. Grading the strength of a body of evidence when assessing health care interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: an update. In: *Methods guide for effectiveness*

- and comparative effectiveness reviews. Rockville, MD: Agency for Healthcare Research and Quality; 2008.
27. Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, Atkins D, Kunz R, Montori V, Jaeschke R, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol* 2013;66(2):151–7.
 28. Eden J, Levit L, Berg A, Morton S, editors. Finding what works in health care: standards for systematic reviews. Washington (DC): National Academies Press (US); 2011.
 29. Nutrition Evidence Library. A series of systematic reviews on the relationship between dietary patterns and health outcomes. Alexandria, VA: Center for Nutrition Policy and Promotion, USDA; 2014.