

The AHRQ National Guideline Clearinghouse (NGC, guideline.gov) Web site will not be available after July 16, 2018 because federal funding through AHRQ will no longer be available to support the NGC as of that date. For additional information, read our [full announcement](#).

EXPERT COMMENTARY DECEMBER 07, 2009

The Many Facets of Evidence: Commentary on the American Academy of Neurology's Evidence Classification Scheme

By: Jacqueline French, MD, FAAN

<https://www.guideline.gov/expert/expert-commentary/164>

Go

APR JUN JUL

09



7 captures

30 Sep 2016 - 12 Jul 2018

2016 2018 2019

About this capture

It often comes as a surprise to individuals who are end users of guidelines that each organization approaches guidelines with a unique perspective. This perspective evolves over time and may result in guidelines from the same organization, written at various times, that follow separate sets of rules. The American Academy of Neurology (AAN) has produced a large number of evidence-based documents over several decades. To ensure consistency, the AAN utilizes a uniform approach to development. For example, its evidence-based documents present clinicians with an outline of available information on questions related to specific diseases or therapeutic interventions and are classified based on the strength of available evidence derived from research (1, 2). Different and distinct classification schemes are used to address screening, diagnostic, prognostic, and therapeutic questions. The documents are called practice parameters and technology assessments rather than guidelines to emphasize that they are based purely on evidence and do not include "consensus" guidance. In essence, this means that readers are presented with "just the facts," along with recommendations that stem strictly from the data.

Recently, the AAN added sections called "clinical context" to the evidence-based documents. These sections present the evidence in a format designed to assist clinicians. To illustrate how this would work, let's say that a particular disease entity can be treated with three different therapeutic modalities. The first, a pharmacologic intervention, was studied in a high-quality, randomized placebo-controlled trial. The second, also a pharmacologic agent, although commonly used, was introduced and approved by regulatory agencies before controlled trials were the norm, and, thus, the evidence for its use may be scant. The third modality, a surgical intervention, was subjected to a clinical trial, but the trial was not placebo-controlled or blinded due to the difficulty of including these trial elements in a surgical study. Thus, the resulting recommendation will indicate that there is good evidence

The AAN classifies evidence using a four-tiered system (Class I through Class IV), with Class I indicating the strongest evidence and Class IV the weakest (1) (see [Table 1](#) below). The class of evidence is determined both by the rigor of the study and the risk of bias. The criteria used to determine the rigor of the study are similar to those used by other entities that produce guidelines and evidence statements. As with other classification schemes, the presence or absence of randomization and a control group are fundamental elements (3).

Notably, random error does not contribute to determination of study strength. Thus, a very small, underpowered study can be designated Class I. In this case, it is expected that a significant or non-significant p-value, or, in the case of an equivalence or non-inferiority trial, the width of confidence intervals (indicating whether a substantial difference has been ruled out) will provide an indication of whether the study will contribute to the recommendation.

It is important to recognize that the evidence will be classified based on clinical questions that are crafted by the practice parameter authors in advance of literature review. Therapeutic questions are laid out in a very formulaic way, as follows: For <patient population A> does <intervention B> compared to <intervention C> improve <outcome D> (1). This format is equivalent to a PICO question. PICO is an acronym for "Patient and question, Intervention, Comparison, Outcome." Formulating questions in this manner also better matches literature search strategies from search engines such as Medline. Questions may focus on one of several issues addressed by a given research study (for example, an analysis of a subpopulation) or may address several different questions which could pertain to a single study. The same study could provide a high level of evidence related to one question, but inadequate evidence related to another. For example, in the guideline on adverse pregnancy outcomes in women with epilepsy, one study was rated Class III in relation to a question about apgar scores in children born to women with epilepsy, which was considered to be a non-objective outcome, but Class II in relation to frequency of perinatal death, which was objective (see below).

Following are brief discussions of the main elements included in the AAN therapeutic classification scheme.

Comparison (Control) Group and Masking

Control groups (individuals who did not receive treatment, received a sham or placebo treatment, or received alternative treatment) are considered a key element in eliminating bias. Concurrent control is considered stronger than historical control (a pre-post design in which the subject acts as their own control) and is necessary for a study to be graded Class I or II.

death).

Randomization and Confounding

In order to prevent confounding by differences between patients that could account for different outcomes, the strongest studies will employ randomized allocation to treatment. Class I designation requires randomized allocation that is properly performed to prevent any possibility of manipulation, as well as a presentation of the major potential patient differences (age, severity of illness, gender, etc.), thus demonstrating the effectiveness of randomization in balancing confounding factors. If imbalance is present, in the absence of statistical adjustment, the study will be downgraded to Class II. Class II can also be achieved by using a matched cohort design.

Drop-outs and Cross-overs

Studies in which more than 20% of patients drop out, or, in the case of an active controlled trial, large numbers of patients cross over to the alternative treatment, are downgraded by one level, since dropouts and crossovers can occur for non-random reasons. In most cases, studies in which the outcome is not analyzed using intention-to-treat statistical methods are also downgraded. However, equivalence or non-inferiority trials must be analyzed using both intention-to-treat and a completer analysis, and if substantial differences exist between the two, the study may be downgraded.

Other Design Elements

Studies in which a primary outcome variable is not selected in advance, or in which a list of patient inclusion and exclusion criteria are not provided, are downgraded by one level.

Active Control Equivalence Trials

Recently, the AAN developed special classification criteria for active control "equivalence" or "non-inferiority" trials. These criteria can be found in [Table 1](#), under the description of Class I requirements.

Use of the Scheme in Developing Recommendations

The following example taken from the recently published "Practice Parameter update: Management issues for women with epilepsy—focus on pregnancy (an evidence-based review): Vitamin K, folic acid, blood levels, and breastfeeding" illustrates the AAN process starting with the clinical question and concluding with the recommendation and clinical

Question: Does preconceptional folic acid supplementation reduce the risk of birth defects in neonates of WWE taking AEDs?

Analysis of Evidence: To be included in the analysis, articles had to measure the association between preconceptional folic acid use and the outcome of major congenital malformations (MCMs). MCMs were defined as structural abnormalities with surgical, medical, or cosmetic importance. The development of an MCM was considered an objective outcome.

Eleven articles relevant to this question were identified by the literature search. They were rated according to the AAN classification of therapeutic evidence scheme. Six studies were graded Class IV and are not discussed further. The remaining studies were rated Class III.

Among the five Class III articles, one study (n=156) showed an increased risk of MCMs with lack of folic acid supplementation (odds ratio [OR] 16.88, 95% confidence interval [CI] 4.79-59.52). The folic acid supplementation dose in this study was reported as 2.5-5 mg per day. A second Class III study measured a significant association between serum folic acid concentrations <4.4 nmol/L and neonatal malformation (adjusted OR 5.8, 95% CI 1.3-27, p =0.02).

Several Class III studies failed to show an association between folic acid and MCMs but were insufficiently sensitive to exclude a substantial risk reduction from folic acid supplementation. One study reported an OR of 1.67 for MCMs in the offspring of mothers on valproate (VPA) who were not taking folic acid supplementation. However, the result was not significant (95% CI 0.62-4.50). Another study showed no effect of folic acid supplementation (OR 0.86, 95% CI 0.34-2.15), but lacked the statistical precision to exclude a potential benefit. The final study was inconclusive since all WWE who had offspring with MCMs had folic acid supplementation.

Conclusion

The risk of MCMs in the offspring of WWE is possibly decreased by folic acid supplementation (two adequately sensitive Class III studies).

Recommendation

Preconceptional folic acid supplementation in WWE may be considered to reduce the risk of MCMs (Level C).

Clinical Context

Folic acid supplementation is generally recommended to reduce the risk of MCMs during pregnancy, and although the data are insufficient to show that it is effective in WWE, there is

epilepsy, be supplemented with at least 0.4 mg of folic acid daily prior to conception and during pregnancy. There was insufficient published information to address the dosing of folic acid and whether higher doses offer greater protective benefit to WWE taking AEDs.

Summary

The AAN scheme for evidence classification has been crafted over several decades. By maintaining the most stringent, evidence-based format, bias is kept to a minimum, although it can never be eliminated entirely. The downside of this approach is that gaps in the evidence base are very apparent to clinicians because they are not filled in by consensus statements. Thus, guidance and recommendations cannot be provided for all therapies commonly employed in practice.

Table 1: Classification Scheme Requirements for Active Control Equivalence Trials

Class I. A randomized, controlled clinical trial of the intervention of interest with masked or objective outcome assessment in a representative population. Relevant baseline characteristics are presented and substantially equivalent among treatment groups or there is appropriate statistical adjustment for differences.

The following are also required:

- a. Concealed allocation
- b. Primary outcome(s) clearly defined
- c. Exclusion/inclusion criteria clearly defined
- d. Adequate accounting for drop-outs (with at least 80% of enrolled subjects completing the study) and cross-overs with numbers sufficiently low to have minimal potential for bias
- e. For non-inferiority or equivalence trials claiming to prove efficacy for one or both drugs, the following are also required*:
 - 1. The authors explicitly state the clinically meaningful difference to be excluded by defining the threshold for equivalence or non-inferiority.
 - 2. The standard treatment used in the study is substantially similar to that used in previous studies establishing efficacy of the standard treatment (e.g., for a drug, the mode of administration, dose and dosage adjustments are similar to those previously shown to be effective).
 - 3. The inclusion and exclusion criteria for patient selection and the outcomes of patients on the standard treatment are comparable to those

Class II. A randomized, controlled clinical trial of the intervention of interest in a representative population with masked or objective outcome assessment that lacks one criteria a-e above or a prospective matched cohort study with masked or objective outcome assessment in a representative population that meets b-e above. Relevant baseline characteristics are presented and substantially equivalent among treatment groups or there is appropriate statistical adjustment for differences.

Class III. All other controlled trials (including well-defined natural history controls or patients serving as their own controls) in a representative population, where outcome is independently assessed, or independently derived by objective outcome measurement.

Class IV. Studies not meeting Class I, II or III criteria including consensus or expert opinion.

*Note that numbers 1-4 in Class Ie are required for Class II in equivalence trials. If any one of the four is missing, the class is automatically downgraded to a Class III.

Author

Jacqueline French, MD, FAAN
New York, NY

Disclaimer

The views and opinions expressed are those of the author and do not necessarily state or reflect those of the National Guideline Clearinghouse™ (NGC), the Agency for Healthcare Research and Quality (AHRQ), or its contractor, ECRI Institute.

Potential Conflicts of Interest

Dr. French holds personal financial interests in Jazz, Eisai, Valeant, Marinus, Pfizer, and UCB. She has received research funding from the Epilepsy Study Consortium, Epilepsy Therapy Development Project, FACES, UCB, Eisai, Johnson & Johnson, and Merck. Dr. French estimates that 30% of her time is spent in outpatient epilepsy practice.

In addition, Dr French discloses the following affiliations/consulting:

- The International League Against Epilepsy
- The American Society of Experimental Neurotherapeutics

- The Epilepsy Study Consortium
- The Epilepsy Foundation
- GlaxoSmithKline
- Pfizer
- UCB Pharma
- Johnson & Johnson
- Cyberonics
- Ortho-McNeil
- Eisai
- Jazz Pharmaceuticals
- Ovation Pharmaceuticals
- Bial Pharmaceuticals
- Neurovista
- Valeant Pharmaceuticals
- Icagen
- Supernus
- Intranasal
- SK Pharmaceuticals
- Taro Pharmaceuticals
- Neurotherapeutics
- Sepracor
- Novartis

References

1. French J, Gronseth G. Lost in a jungle of evidence: we need a compass. *Neurology* 2008;71:1634-1638.
2. Gronseth G, French J. Practice parameters and technology assessments: what they are, what they are not, and why you should care. *Neurology* 2008;71:1639-1643.
3. Lohr KN. Rating the strength of scientific evidence: relevance for quality improvement programs. *Int J Qual Health Care* 2004;16:9-18.