**Levels of Evidence**

**JEVDIP Improves Grading System and Adopts Strength of Recommendation Taxonomy Grading (SORT) for Guidelines and Systematic Reviews***

**TRANSLATION OF EB RECOMMENDATIONS**

The *Journal of Evidence-Based Dental Practice* (JEVDIP) is adopting the SORT as its method to better evaluate guideline recommendations and the results from systematic reviews1 (Fig. 1).

The quality rating scores for individual studies also change to a simpler 3-tier grade that deemphasizes observational studies because the evidence from observational studies is normally too weak to support clinical recommendations.

For some years, the standard approach to evaluate the quality of individual studies was based on a hierarchical grading system of research design in which randomized control trials (RCTs) received the highest score. The JEVDIP has published many articles and evaluations using these methods. For single clinical trials, the quality grade is still very useful and JEVDIP will continue to provide a grade of its analysis of these publications (see later in this article).

The widespread acceptance of good critical appraisal techniques has also drawn attention to the limitations of these evaluation methods in providing guidance for clinical decision making. An important reason to move from just a quality grade to strength of recommendation is that there is no consensus on how to assign a weight to each of the individual study’s major and minor defects. Without this, there is large variation in how evaluators assign a grade on a consistent basis. The “new” system to be used by JEVDIP is as follows:

- Of great importance to the JEVDIP editors is our commitment to use a grading scale for the strength of recommendation that could be applied more uniformly by reviewers and readers with varying degrees of expertise in evidence-based techniques and clinical epidemiology.

The SORT “grade” is derived from evaluating the evidence from a body of knowledge most often about a single procedure or new product.1,2 The strength grades are much more useful to the clinician because they emphasize effectiveness and are intended to provide results that are more applicable to “average” patients. The quality of the evidence on effectiveness is a key component, but not the only component, in making good clinical decisions.

While evidence-based approaches for reviewing clinical research have helped to minimize interpretation bias,3 methods for connecting that evaluation with clinical recommendations have recently started to appear.4 Initially, clinical recommendations by evidence-based groups and professional organizations5,6 were strongly correlated with the research design of the most important studies. Now for many subjects, there are larger amounts of information available making it also necessary to consider the evidence in its entirety, lending itself to a strength of recommendation grade.

**STRENGTH OF RECOMMENDATION TAXONOMY (SORT)**

The original authors of SORT represent several of the major family medicine journals in the United States (ie, *American Family Physician, Family Medicine, The Journal of Family Practice, Journal of the American Board of Family Practice*, and *BMJ-USA*) and a large family practice academic consortium (Family Practice Inquiries Network [FPIN]).1

They decided that the proposed taxonomy, or system, for rating the strength of a recommendation should address 3 key elements they identified in an often-quoted Agency for Health Quality Research (AHQR) report.7

- Quality of evidence
- Quantity of evidence
- Consistency of evidence

**THE SYSTEM**

The strength of the recommendation is graded either A, B, or C using the criteria in Table 1.

Recommendations based only on improvements in surrogate or disease-oriented outcomes (efficacy) are always categorized as level C. This is because improvements in efficacy are not always associated with improvements in patient-oriented outcomes (effectiveness).8,9

The 2 factors used to determine the final SORT grade are

1. the quality of the individual studies (Table 2)
2. and the consistency of evidence across all the studies being evaluated (Table 3).

**Figure 1** gives information about how to determine the strength of recommendation grade using an algorithm.

- Reviewers and readers should adjust the strength of recommendation grade based on an individual patient or circumstance based on the benefits, harms, and costs of the intervention being recommended.

**STUDY QUALITY**

The quality of individual studies is rated levels 1, 2, or 3. Table 2 explains how to determine the level of evidence for an individual study. Once the quality grade is known, it can be “plugged in” to the algorithm in Figure 1, along with the consistency grade (Table 3) to yield the final SOR.

**TABLE 1. Strength-of-Recommendation Grades**

<table>
<thead>
<tr>
<th>Strength of recommendation</th>
<th>Basis for recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Consistent, good-quality patient-oriented evidence*</td>
</tr>
<tr>
<td>B</td>
<td>Inconsistent or limited-quality patient-oriented evidence*</td>
</tr>
<tr>
<td>C</td>
<td>Consensus, disease-oriented evidence,* usual practice, expert opinion, or case series for studies of diagnosis, treatment, prevention, or screening</td>
</tr>
</tbody>
</table>

*Patient-oriented evidence measures outcomes that matter to patients: morbidity, mortality, symptom improvement, cost reduction, and quality of life. Disease-oriented evidence measures intermediate, physiologic, or surrogate end points that may or may not reflect improvements in patient outcomes (e.g., blood pressure, blood chemistry, physiologic function, pathologic findings).
grade. Note that studies about etiology are always given a quality grade of 3 since they represent disease-oriented evidence.

**SUMMARY**

There are several advantages to SORT over other systems:

- It is straightforward and comprehensive
- Is easily applied by reviewers and clinicians
- It explicitly addresses the issue of patient-oriented (effectiveness) versus disease-oriented evidence (efficacy)

SORT strengths also create limitations. Some clinicians may be concerned that the system is not as detailed in its assessment of study designs as others, such as the one previously used by JEBDP, which was adapted from the Center for Evidence-Based Medicine (CEBM). However, SORT authors believe the primary difference between the 2 systems is that the CEBM version also distinguishes...
between good and poor observational studies whereas the SORT version does not.

JEBDP editors agree with SORT authors in concluding that the advantages of a system that provides the clinician with a clear recommendation that is strong (A), moderate (B), or weak (C) in its support of a particular intervention outweighs the theoretical benefit of distinguishing between lower quality and higher quality observational studies. JEBDP editors agree with this view because the evidence from observational studies is normally too weak to support clinical recommendations.

Like any such grading scale, SORT is a work in progress and we can expect changes in the future.

REFERENCES


