LEVELS OF EVIDENCE—JOURNAL OF EVIDENCE-BASED DENTAL PRACTICE GRADING SYSTEM*

STRENGTH OF RECOMMENDATION TAXONOMY (SORT)

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 (Figure 1) These SORT 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.

The proposed taxonomy, or system, for rating the strength of a recommendation should address 3 key elements 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 (SORT) is graded either A, B, or C using the criteria in Table 1

<table>
<thead>
<tr>
<th>SORT SCORE</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).

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

- The quality of the individual studies (Table 2)
- The consistency of evidence across all the studies being evaluated (Table 3)

Figure 1. Assigning a Strength-of-Recommendation grade based on a body of evidence. (USPSTF ¼ U.S. Prevention Services Task Force)

Table 2. Assessing Quality of Evidence

<table>
<thead>
<tr>
<th>LEVEL OF EVIDENCE</th>
<th>Diagnosis</th>
<th>Treatment / prevention / screening</th>
<th>Prognosis</th>
</tr>
</thead>
</table>
| 1 good-quality, patient-oriented evidence | • Validated clinical decision rule  
• SR/meta-analysis of high-quality studies  
• High-quality diagnostic cohort study* | • SR/meta-analysis or RCTs with consistent findings  
• High-quality individual RCT†  
• All-or-none study‡ | • SR/meta-analysis of good-quality cohort studies  
• Prospective cohort study with good follow-up |
| 2 limited-quality patient-oriented evidence | • Unvalidated clinical decision rule  
• SR/meta-analysis of lower quality studies or studies with inconsistent findings  
• Lower quality diagnostic cohort study or diagnostic case-control study | • SR/meta-analysis of lower quality clinical trials or of studies with inconsistent findings  
• Lower quality clinical trial  
• Cohort study  
• Case-control study | • SR/meta-analysis of lower quality cohort studies or with inconsistent results  
• Retrospective cohort study or prospective cohort study with poor follow-up  
• Case-control study  
• Case series |
| 3 other evidence | Consensus guidelines, extrapolations from bench research, usual practice, opinion, disease-oriented evidence (intermediate or physiologic outcomes only), or case series for studies of diagnosis, treatment, prevention, or screening | | |

SR, systematic review; RCT, randomized controlled trial.

* High-quality diagnostic cohort study: cohort design, adequate size, adequate spectrum of patients, blinding, and a consistent, well-defined reference standard.
† High-quality RCT: allocation concealed, blinding if possible, intention-to-treat analysis, adequate statistical power, adequate follow-up (greater than 80 percent).
‡ In an all-or-none study, the treatment causes a dramatic change in outcomes, such as antibiotics for meningitis or surgery for appendicitis, which precludes study in a controlled trial.
**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 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 grade. Note that studies about etiology are always given a quality grade of 3 since they represent disease-oriented evidence.

<table>
<thead>
<tr>
<th>Consistent</th>
<th>Inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most studies found similar or at least coherent conclusions (coherence means that differences are explainable). or If high-quality and up-to-date systematic reviews or meta-analyses exist, they support the recommendation.</td>
<td>Considerable variation among study findings and lack of coherence or If high-quality and up-to-date systematic reviews or meta-analyses exist, they do not find consistent evidence in favor of the recommendation.</td>
</tr>
</tbody>
</table>

**SUMMARY**

There are several advantages to SORT over other systems:

- It is straightforward and comprehensive
- It 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**


