# APPENDIX B

## KP System for Grading the Strength of a Body of Evidence

<table>
<thead>
<tr>
<th>Level/Grade</th>
<th>Therapy/Prevention/Screening</th>
<th>Diagnosis</th>
<th>Prognosis/Etiology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GOOD</strong></td>
<td><strong>Type and number of studies</strong></td>
<td><strong>Type and number of studies</strong></td>
<td><strong>Type and number of studies</strong></td>
</tr>
<tr>
<td></td>
<td>At least one well-designed and conducted systematic review (SR/MA) (consider heterogeneity) of RCTs</td>
<td>At least one well-designed and conducted SR/MA (consider heterogeneity) of cross-sectional studies using independent gold standard</td>
<td>At least one well-designed and conducted SR/MA (consider heterogeneity) of prospective cohort studies</td>
</tr>
<tr>
<td></td>
<td>Two or more well-designed and conducted RCTs with narrow confidence intervals</td>
<td>Two or more well-designed and conducted cross-sectional studies using an independent gold standard</td>
<td>Two or more well-designed and conducted prospective cohort studies</td>
</tr>
<tr>
<td></td>
<td>One well-designed and conducted multicenter RCT with narrow confidence intervals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality</td>
<td>Low risk of bias</td>
<td>Low risk of (verification) bias</td>
<td>Low risk of bias</td>
</tr>
<tr>
<td></td>
<td>Adequate sample size and power</td>
<td>Independent gold standard</td>
<td>Acceptable loss to follow-up</td>
</tr>
<tr>
<td></td>
<td>No major methodological concerns</td>
<td>No major methodological concerns</td>
<td>No major methodological concerns</td>
</tr>
<tr>
<td>Consistency</td>
<td>For SR/MA, no major conflict in results (consider heterogeneity). If significant heterogeneity exists, drops to Poor</td>
<td>For SR/MA no major conflict in results (consider heterogeneity)</td>
<td>For SR/MA no major conflict in results (consider heterogeneity)</td>
</tr>
<tr>
<td></td>
<td>For individual RCTs, no major conflict in results</td>
<td>For individual studies, consistent diagnostic accuracy</td>
<td>For SR/MA no major conflict in results (consider heterogeneity)</td>
</tr>
<tr>
<td></td>
<td>If major conflicts do exist, drop to “Insufficient”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relevancy</td>
<td>No compelling reason not to generalize published work to target KP population</td>
<td>No compelling reason not to generalize published work to target KP population</td>
<td>No compelling reason not to generalize published work to target KP population</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FAIR</th>
<th><strong>Type and number of studies</strong></th>
<th><strong>Type and number of studies</strong></th>
<th><strong>Type and number of studies</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single well-designed and conducted RCT with narrow confidence intervals</td>
<td>Single well-designed and conducted cross-sectional study</td>
<td>Single well-designed and conducted prospective cohort study</td>
</tr>
<tr>
<td></td>
<td>Two or more RCTs of lower quality</td>
<td>Two or more cross-sectional studies of lower quality</td>
<td>Two or more well-designed, well-conducted retrospective cohort studies</td>
</tr>
<tr>
<td></td>
<td>Well-designed and conducted SR/MA of cohort studies (consider heterogeneity)</td>
<td>Well-designed and conducted SR/MA of lower quality studies</td>
<td>Well-designed and conducted SR/MA of lower quality studies</td>
</tr>
<tr>
<td>For screening interventions only, the following are also acceptable as Fair evidence:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Two or more well-designed and conducted cohort studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Two or more well-designed and conducted case-control studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Two or more well-designed and conducted case-control studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality</td>
<td>Minor methodological concerns</td>
<td>Minor methodological concerns</td>
<td>Minor methodological concerns</td>
</tr>
<tr>
<td>Consistency</td>
<td>For SR/MA, no major conflict in results (consider heterogeneity)</td>
<td>For SR/MA, no major conflict in results (consider heterogeneity)</td>
<td>For SR/MA, no major conflict in results (consider heterogeneity)</td>
</tr>
<tr>
<td></td>
<td>For individual studies, no major conflict in results</td>
<td>For individual studies, no major conflict in results</td>
<td>For individual studies, no major conflict in results</td>
</tr>
<tr>
<td></td>
<td>If major conflicts do exist, drop to “Insufficient”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relevancy</td>
<td>No compelling reason not to generalize published work to target KP population</td>
<td>No compelling reason not to generalize published work to target KP population</td>
<td>No compelling reason not to generalize published work to target KP population</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INSUFFICIENT</th>
<th><strong>Type and number of studies</strong></th>
<th><strong>Type and number of studies</strong></th>
<th><strong>Type and number of studies</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single RCT of lower quality or insufficient size</td>
<td>Single cross-sectional study of lower quality</td>
<td>Single prospective cohort study of lower quality</td>
</tr>
<tr>
<td></td>
<td>Cohort study</td>
<td>Case-control study</td>
<td>Retrospective cohort study</td>
</tr>
<tr>
<td>Quality</td>
<td>Major methodological concerns (i.e., lack of concealed allocation, inadequate blinded, no ITT analysis)</td>
<td>Major methodological concerns (non-consecutive, poor or non-independent gold standard)</td>
<td>Major design or methodological concerns</td>
</tr>
<tr>
<td>Consistency</td>
<td>Studies that are well-designed and conducted (Good or Fair) but with major conflict in results</td>
<td>Studies that are well-designed and conducted (Good or Fair) but with major conflict in results</td>
<td>Studies that are well-designed and conducted (Good or Fair) but with major conflict in results</td>
</tr>
<tr>
<td></td>
<td>SR/MA with major conflict in results (consider heterogeneity)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relevancy</td>
<td>Compelling reasons why the results do not apply to the target KP population</td>
<td>Compelling reasons why the results do not apply to the target KP population</td>
<td>Compelling reasons why the results do not apply to the target KP population</td>
</tr>
</tbody>
</table>

*Evidence is graded with respect to the degree it supports the specific clinical recommendation. For example, there may be good evidence that Drugs 1 and 2 are effective for Condition A, but no evidence that Drug 1 is more effective than Drug 2. If the recommendation is to use either Drug 1 or 2, the evidence is good. If the recommendation is to use Drug 1 in preference to Drug 2, the evidence is insufficient.*