Application of GRADE to Cochrane Diagnostic Test Accuracy Reviews

Gopalakrishna G1, Mustafa R2, Langendam M3, Leeflang M1, Bossuyt P1

1Department of Clinical Epidemiology, Academic Medical Centre, University of Amsterdam
2Department of Clinical Epidemiology and Biostatistics, McMaster University, Canada
3Dutch Cochrane Centre, Academic Medical Centre, University of Amsterdam
www.decide-collaboration.eu/WP4

*Contact person: g.gopalakrishna@amc.uva.nl

BACKGROUND
The GRADE Working Group has a rigorous methodology for rating the evidence and making recommendations for interventions. This methodology can be used for evaluating diagnostic accuracy studies, but experience is limited and the method under development (1).

AIM
To identify challenges in the GRADE methodology when used to rate the quality of evidence from published diagnostic test accuracy reviews (DTAR).

METHODS
We selected three Cochrane DTARs based on diversity of clinical areas and methodological issues. These were reviews by Virgili et al 2011 (Cochrane Eyes and Vision Group), van der Windt D et al 2010 (Cochrane Back Group) and Abba K et al 2012 (Cochrane Infectious Diseases Group). At least 2 reviewers rated the evidence according to the five “GRADE domains” as summarised below. Assessors explained judgments made on the quality of the evidence by documenting all considerations. Two teleconferences were held to discuss the issues faced when applying the GRADE methodology to DTARs.

RESULTS: Summary of the main issues

<table>
<thead>
<tr>
<th>GRADE DOMAINS</th>
<th>KEY ISSUES / CONSIDERATIONS ENCOUNTERED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk of Bias/ Study Limitations</td>
<td>Assessors were unclear on how to judge QUADAS items labeled “unclear”</td>
</tr>
<tr>
<td>Indirectness</td>
<td>i. In at least one review, review authors did not link the index test to a care pathway ii. Test accuracy is inherently indirect evidence for patient outcomes, would this then warrant a default downgrading of the quality of the evidence,</td>
</tr>
<tr>
<td>Inconsistency</td>
<td>Assessors used different rationales for downgrading the evidence (e.g. CI overlap, unexplained heterogeneity, inconsistent use of test threshold positivity and variable ref std definitions)</td>
</tr>
<tr>
<td>Imprecision</td>
<td>Assessors used different rationales for downgrading the evidence (e.g. small study no., wide CI)</td>
</tr>
<tr>
<td>Publication bias</td>
<td>Assessors were unclear on how to assess this</td>
</tr>
</tbody>
</table>

ACROSS ALL DOMAINS

Assessors had to be conscious to not double downgrade on a single factor e.g. reviews with small study numbers could be downgraded under “imprecision” (due to wide CI) or under “Risk of Bias / Study Limitations”

FOR COMPARATIVE TEST REVIEWS

i. Each test was assessed first against its ref standard and then relative to each other. Some assessors thus felt the need to create three separate tables ii. When making the relative comparison, the score for each GRADE domain was determined as the lower of the 2 scores for that domain for each index test when compared to its ref standard iii. Overall quality of evidence be further downgraded by one level for indirectness

GENERAL ISSUES: CLINICAL QUESTION

Clear PICO styled key question important esp in DTARs comparing multiple index tests or with different patient spectrums

CONCLUSIONS
1. Clear definitions of the GRADE domains “inconsistency”, “imprecision” and “publication bias” with respect to DTARs would facilitate the operationalization of GRADE for diagnostics
2. Explicit guidance on how to rate the quality of evidence for a comparative test review is needed.

References: Schünemann H et al. 2008. BMJ 336; 1106-10