

## 074 IMPLEMENTATION OF GRADE METHODOLOGY IN RECOMMENDATION DEVELOPMENT

<sup>1</sup>J Haynes, <sup>1</sup>D Regidor, <sup>2</sup>W Chan. <sup>1</sup>Kaiser Permanente, Care Management Institute, Oakland, USA; <sup>2</sup>Kaiser Permanente, NW Permanente Group, Oregon, USA

10:1136/bmjqs-2013-002293.105

**Background** We have been developing CPGs for use within our organisation since 2002. Our lengthy, text-based rationales were not widely read by guideline users. We created a decision support (rationale) table, based on GRADE methodology, and added a summary statement (basis of recommendation) to allow readers a concise and transparent snapshot of our justification for recommendation and strength.

**Context** The rationale serves as a bridge between systematic review and recommendation, and provides users with a high-level justification for a recommendation. The basis of recommendation (BoR) summarises the 4 GRADE domains of strength of recommendation and how they are integrated to derive the final recommendation and strength. The BoR serves to: •Provide information to the Guideline Development Team and frontline clinicians to facilitate discussion, consensus and aid clinical decision-making. •Provide a structured, standardised portal into more detailed information in the CPG.

**Description of Best Practice** We follow GRADE's 2-level designation of recommendation strength (strong/weak), and developed standardised recommendation language to align with recommendation strength. We considered two approaches to derive the final recommendation strength, finally settling on an approach that allows flexible weighting of the contribution of each domain to recommendation strength. With this approach, in special circumstances, a strong recommendation may be given in the absence of a high-level of certainty. We plan to provide direct links from the CPG to our electronic medical record's decision support tools.

**Lessons for Guideline Developers, Adaptors, Implementers, and/or Users** A concise and targeted rationale helps clinicians understand how evidence is used to develop clinical practice recommendations.

## 075 THE USE OF GRADE METHODS IN THE WORLD HEALTH ORGANIZATION (WHO) PUBLIC HEALTH GUIDELINES (PHG): DISTRIBUTION OF STRENGTH OF RECOMMENDATIONS AND CONFIDENCE IN ESTIMATES OF EFFECT

<sup>1</sup>P Elias, <sup>2</sup>S Norris, <sup>3</sup>J Brito, <sup>4</sup>R Stoltzfus, <sup>5</sup>L Bero, <sup>6</sup>B Djulbegovic, <sup>1,7</sup>I Neumann, <sup>3</sup>V Montori, <sup>1</sup>G Guyatt. <sup>1</sup>McMaster University, Hamilton, Canada; <sup>2</sup>World Health Organization, Geneva, Switzerland; <sup>3</sup>Mayo Clinic, Rochester, MN, USA; <sup>4</sup>Cornell University, Ithaca, USA; <sup>5</sup>University of California, San Francisco, USA; <sup>6</sup>University of South Florida, Tampa, USA; <sup>7</sup>Pontificia Universidad Catolica de Chile, Santiago, Chile

10:1136/bmjqs-2013-002293.106

**Background** A perception exists that expert guideline panellists are sometimes reluctant to offer weak/conditional/contingent recommendations. GRADE guidance warns against strong recommendations when confidence in estimates of effect is low or very low (low or very low quality evidence), suggesting that such recommendations may seldom be justified.

**Objectives** To characterise the distribution of strength of recommendations and confidence in estimates of effect in WHO guidelines that have used the GRADE approach and graded strength of recommendations and confidence in effect estimates.

**Methods** We reviewed guidelines listed in the WHO guidelines database as of November 2012. We identified those that use GRADE and, in these guidelines, examined the distributions of strong and weak and associated confidence in estimates of effect (high, moderate, low, and very low).

**Results** We identified 116 WHO guidelines; 48 (41.3%) referred to GRADE methods, and 43 (37%) utilised GRADE and provided both a strength of recommendation and confidence in estimates grading. These 43 guidelines included 456 recommendations, of which 290 (63.6%) were strong and 166 (36.4%) were conditional/weak. Of the 290 strong recommendations, 97 (33.4%) were based on evidence warranting low confidence in estimates of effect and 63 (21.7%) on evidence warranting only very low confidence.

**Discussion** Strong recommendations based on low or very low confidence in effect estimates are very frequently made in WHO guidelines. Further study to determine the reasons for such recommendations is warranted.

**Implications for Guideline Developers/Users** Guideline authors should provide a clear, compelling rationale for any strong recommendations based on low confidence estimates.

## 076 ENHANCING THE ACCEPTANCE AND IMPLEMENTATION OF GRADE SUMMARY TABLES FOR EVIDENCE ABOUT DIAGNOSTIC TESTS

<sup>1,2</sup>R Mustafa, <sup>1</sup>W Wiercioch, <sup>1</sup>J Brozek, <sup>3</sup>M Lelgemann, <sup>3</sup>D Buehler, <sup>1,4</sup>A Garg, <sup>5</sup>P Bossuyt, <sup>1</sup>H Schünemann. <sup>1</sup>McMaster University, Hamilton, Canada; <sup>2</sup>University of Missouri-Kansas City, Kansas City, USA; <sup>3</sup>GKV, Berlin, Germany; <sup>4</sup>University of Western Ontario, London, Canada; <sup>5</sup>University of Amsterdam, Amsterdam, The Netherlands

10:1136/bmjqs-2013-002293.107

**Background** The GRADE Working Group developed Summary tables adapted to summarise and present evidence from diagnostic test accuracy (DTA) systematic reviews.

**Objective** To develop guidance on what information to include in these summary tables and to determine the best method(s) for presentation for different end users, including healthcare providers, systematic reviewers and guideline developers.

**Methods** We presented a number of alternative summary tables to participants. We conducted questionnaires and one-on-one user testing interviews with target end users. We presented printed copies of summary tables and asked open-ended and 7-point Likert-scale questions to obtain information about users' understanding and preferences.

**Results** All participants (n=60) agreed that using summary tables to present results of DTA reviews is helpful. Presentation of several disease prevalence values was identified as a source of confusion. There was an overall preference for placement of sensitivity and specificity values inside summary tables to allow making a link to individual test results (TP, FN, TN, FP). A third of the participants read explanatory content in table footnotes. Two thirds of the participants noted that additional data, including adverse effects, costs, and treatment consequences, would be helpful for making appropriate conclusions and decisions about diagnostic tests.

**Discussion** As results of DTA reviews are conceptually complicated, presenting the data in a clear, comprehensive, comprehensible way that is tailored to different end users is critical.

**Implications for Guideline Developers/Users** We are developing a 3-layer approach, with varied content in summary tables of each layer tailored to the needs of different end users.