

The recent IOM standards were developed by as best practice with little guidance for potential interpretation or resource requirements. The CMSS Principles are intended to be a step towards practical guidelines standards and this research the first feedback step as to measuring that practicality.

P266 AT WHAT RATE DOES NEW EVIDENCE CHANGE GUIDANCE

¹J Ording, ²M Thiese, ²K Hegmann, ¹C Wolfkiel, ²U Ott. ¹American College of Occupational and Environmental Medicine, Elk Grove Village, USA; ²University of Utah RMCOEH, Salt Lake City, USA

10:1136/bmjqs-2013-002293.229

Background New research is continually shaping guidelines; however, the rate of change has not been assessed.

Objectives Review articles from 2010 to 2013 to assess rate of change of guidelines for the elbow.

Methods A systematic literature search was conducted to identify randomised controlled trials (RCTs) on elbow disorders published between 2010 and 2013. Identified RCTs were scored using established scoring methods and incorporated into guidelines to determine if any recommendations needed to be changed or new recommendations added.

Results Fifteen new RCTs were identified (4 high-, 9 moderate-, and 2 low-quality). Nine (69%) studies were used to make 3 new recommendations and 4 changes to recommendations to guidelines on elbow disorders. Seven of these studies prompted new guidance on soft tissue mobilisation, autologous blood injections, periarticular blood injections for lateral epicondylalgia (LE). Two of these studies caused changes to the recommendation level for manipulation/mobilisation for LE and evidence level changes for exercises, glucocorticosteroid injections, and platelet rich plasma injections for LE. Seven (53%) studies did not change any of the recommendations but added to the body of evidence to support the current recommendations.

Discussion New studies may be higher quality and have significant impact on guidelines. Two-thirds of new evidence triggered recommendations changes or development of new recommendations for treating LE. Additional assessments of low back and other body parts are underway.

Implications for Guideline Developers/Users It is beneficial to do a yearly review of the literature to determine if any new evidence will impact changes to current guidelines.

P268 WHAT KIND OF EVIDENCE SUPPORTED THE CLINICAL PRACTICE GUIDELINE FOR THE SYNDROMIC MANAGEMENT OF SEXUALLY TRANSMITTED INFECTIONS AND OTHER INFECTION OF THE GENITAL TRACT 2012

^{1,2}H Gaitan, ^{1,2}M Torres, ¹A Rodriguez. ¹Clinical Research Institute, Universidad Nacional de Colombia, Bogota, Colombia; ²Sexually Transmitted Infections Cochrane Group, Bogota, Colombia

10:1136/bmjqs-2013-002293.230

Background In recent years, the GRADE approach has been broadly accepted by many GDG. There is a lack of information about the relationship between the types of evidence with the strength of recommendations using GRADE.

Objective To present the evidence mapping of the literature used to develop recommendations in the Guideline for Syndromic management of the Genital Tract Infections.

Methods The evidence of each recommendation was reviewed and was classified according to the type, quality and quantity of evidence and strength of the recommendation. A descriptive analysis was performed as well as a cross-analysis to find out the relationship between the strength of the recommendation and the underlying quality of the evidence.

Results 80 recommendations were identified. Systematic reviews supported the 29.1% of the recommendations, RCT 25.6%, observational studies 5.9%, guidelines 9.3% and expert opinion 30.1%. The quality of the evidence was high (14%), moderate (15%), low (16%) and very low (55%). 63.7% of the recommendations were strong in favour. 14% of the strong recommendations came from high quality evidence and 49% came from very low quality evidence.

Discussion The evidence shows a similar percentage of systematic reviews, RCT and expert opinion in the guideline. Despite the quality of the evidence, the number of strong recommendations is high due to the other criteria of the GRADE approach.

Implications for Guideline Developers The GRADE approach allows weighting other factors beyond the quality of the evidence. Research needs to be done on the most important factors in grading the recommendations.

P269 ADAPTATION OF A NORTH AMERICAN INSTITUTIONALLY BASED HEALTH TECHNOLOGY ASSESSMENT (IHATA) MODEL TO A PRIVATE BRAZILIAN HEALTH CARE ORGANIZATION (BHCO)

^{1,2}S Alves Da Silva, ¹M Cabanelas Pazos, ¹P Pereira de Souza, ¹P Rascão Cardoso, ^{2,3}P Wyer. ¹Amil Assistência Médica Internacional, Rio de Janeiro, Brazil; ²The New York Academy of Medicine, New York, USA; ³Columbia University, New York, USA

10:1136/bmjqs-2013-002293.231

Background Brazilian Guidelines, developed by medical societies, are sparsely used by federal agencies to determine coverage. To date there is no organised approach to clinical policy and guideline development or for dialogue with regulators within BHCOs. Amil, the largest BHCO, covers 6 million lives distributed across 8 regions and delivers care to many through its own medical centres.

Objectives To develop a minimally resourced clinical policy and implementation capability within Amil together with a training programme on a national level.

Methods Our approach is based on observation of the Penn Health System and the Kaiser Permanente (KP) models of IHATA. We are profiling current capacity for integrating umbrella reviews with mining and interpretation of internally generated practice data, and are identifying resource and manpower needs. To promote cultural change on national scale we reformulated an annual training workshop made in partnership with NYAM and McMaster and opened to participants within and without Amil, including Health Ministry and Regulatory Agencies, by addressing guideline development, adaptation and implementation skills.

Results Our approach identifies knowledge gaps within the organisation and develops related guidelines and outcomes assessment to be internally used through Electronic Health Records and to be presented to regulators as proposal for change. The framework was built on a piloted approach on

bariatric surgery and on the identification of additional target areas. Our capacity building programme has trained over 250 individuals within and without Amil, including policy developers within regional and national health ministries.

Conclusion An IHTA programme within a BHCO may offer a capacity building model useful for national development.

P270 DEVELOPMENT OF QUALITY INDICATORS OF LOW RISK LABOR PROVIDED PRIMARILY BY MIDWIVES: BASED ON CLINICAL PRACTICE GUIDELINES WITH A MODIFIED DELPHI METHOD

^{1,2}K Ueda, ¹S Ohtera, ^{1,3}K Misato, ¹T Nakayama. ¹Department of Health Informatics, Kyoto University School of Public Health, Kyoto, Japan; ²Morinomiya University of Medical Sciences School of Nursing, Osaka, Japan; ³Kansai University of Health Sciences School of Nursing, Osaka, Japan

10:1136/bmjqs-2013-002293.232

Background In Japan some labour units have started services for low risk labour provided primarily by midwives with the assistance of obstetricians. However, real practices varied with facilities and quality of care has not been sufficiently assured.

Objectives We aimed to develop quality indicators of low-risk labour provided primarily by midwives in Japan.

Methods We systematically searched databases to select related indicators and clinical practice guidelines. Using a modified Delphi method, a multidisciplinary panel including two women who experienced labour and were not health professionals was assembled and a meeting was held. After the panel members rated the candidates individually, consensus was attained for each indicator.

Results An electronic search retrieved 104 guidelines (2043 key recommendations) from which 32 guidelines (166 key recommendations) were selected. Some recommendations were excluded due to difficulties in measuring and others due to evidence against following certain guidelines. Additionally, 31 existing quality indicators were identified. The panel discussed 25 candidate indicators: 18 indicators were adopted, six indicators were not adopted, one indicator was modified. Four indicators were added later following discussion. Consequently, 23 quality indicators of care were chosen.

Discussion A pilot test of these indicators at multiple facilities should contribute to future development of guideline-based quality indicators.

Implications for Guideline Developers/Users Guideline developers should be more conscious in making recommendations that are potentially used as quality indicators.

P271 COMPUTERIZED GUIDELINES ANALYSIS: 5-YEARS EXPERIENCE AND USER FEEDBACK

¹G Georg, ²J Siddle, ²M Truran, ²M Cavazza. ¹Haute Autorité de Santé, Saint-Denis, France; ²Teesside University, Middlesbrough, UK

10:1136/bmjqs-2013-002293.233

Background In 2007, HAS (Haute Autorité de Santé, France) started experimenting the use of computerised guidelines analysis with the G-DEE software as part of the guidelines production process. G-DEE uses shallow natural language processing techniques to recognise the expression of recommendations in a guidelines' text, and structures the document accordingly using on-

screen text highlighting and colouring. Project managers were offered, on an entirely voluntary basis, the possibility of using this tool during the final phases of guidelines production to i) analyse overall guidelines structure and ii) review the formulation of individual recommendations. To date, this experiment has concerned 38% of guidelines that have been officially released.

Objectives Improving the performance and usability of the deployed G-DEE system.

Methods We have collected feedback from guidelines' project managers who have identified clear benefits in terms of readability of individual recommendations, overall text balance, fine-tuning and testing of alternative wording. Main limitations identified were processing time, which required additional planning and the absence of a tuning facility that would allow different levels of grammar coverage.

Results Notable enhancements include faster document processing (x10) and incremental document analysis allowing direct text input in the system.

Discussion These will make its use possible throughout the production process, e.g. directly as part of consensus meetings once the first draft is available. In addition, the possibility to define additional grammar files for text processing supports the processing of new document types (shorter guidelines, summaries) which may adopt different writing styles.

P273 EFFECTS OF KOREAN-AGREE SCORING GUIDE ON IMPROVING THE RELIABILITY OF THE SCORES

¹M Oh, ^{1,2,5}H Jo, ^{3,4,5}Y Lee. ¹Department of Preventive Medicine, Kangwon National University Hospital, Chuncheon, Korea; ²Department of Health Management and Policy, Kangwon National University, Chuncheon, Korea; ³Department of Laboratory Medicine and Genetics, Soonchunhyang University Hospital, Bucheon, Korea; ⁴Department of Laboratory Medicine and Genetics, Soonchunhyang University, Bucheon, Korea; ⁵The Executive Committee for Clinical Practice Guideline, The Korean Academy, Seoul, Korea

10:1136/bmjqs-2013-002293.234

Background The Korean translated AGREE II instrument has been distributed to Korean medical societies in 2011. But inter-rater disagreement issues still existed. The Korean AGREE II scoring guide was therefore developed to reduce inter-rater differences.

Objectives To examine effects of the Korean AGREE II scoring guide on reduction of the inter-rater differences.

Methods Appraisers were randomly assigned to group A or B. The Korean AGREE II instrument was provided to all participants, but the Korea AGREE II scoring guide was offered to only group A. Total of 14 appraisers participated and each guideline was assessed by 8 appraisers. To evaluate the reliability of the Korean AGREE II scoring guide, correlation of scores among appraisers and domain-specific intra-class correlation (ICC) were calculated.

Results Most scores of two groups were comparable. And scoring guide user group showed higher reliability at all guidelines. It showed higher correlation among appraisers and higher ICC values at almost all domains (from 0.272~0.792 to 0.667~0.888 in CPG A, from 0.273~0.682 to 0.424~0.815 in CPG B).

Discussion Scoring guide reduces the inter-rater disagreement and improves the overall reliability of the Korean-AGREE II instrument.