

**Implications for Guideline Developers/Users** Guideline developers can use GRADE and these methods when there is no evidence or low/very low quality evidence from RCTs.

#### 056 ADDRESSING CONTINUOUS DATA FOR PARTICIPANTS EXCLUDED FROM TRIAL ANALYSIS: A GUIDE

<sup>1</sup>S Ebrahim, <sup>2</sup>E Akl, <sup>3</sup>R Mustafa, <sup>3</sup>X Sun, <sup>1</sup>S Walter, <sup>1</sup>D Heels-Ansdell, <sup>4</sup>P Alonso-Coello, <sup>1</sup>B Johnston, <sup>1</sup>G Guyatt. <sup>1</sup>McMaster University, Hamilton, Canada; <sup>2</sup>American University of Beirut, Beirut, Lebanon; <sup>3</sup>The Centre for Health Research, Kaiser Permanente Northwest, Portland, USA; <sup>4</sup>Iberoamerican Cochrane Centre, CIBERESP-IIB Sant Pau, Barcelona, Spain; <sup>5</sup>University of Missouri, Kansas City, USA

10:1136/bmjqs-2013-002293.87

**Background** Guideline developers addressing quality of evidence commonly confront studies with missing data.

**Objectives** To develop a framework for assessing risk of bias resulting from missing participant data for continuous outcomes in systematic reviews.

**Methods** We developed a range of progressively more stringent imputation strategies to challenge the robustness of the pooled estimates. We applied our approach to two systematic reviews.

**Results** We used 5 sources of data for imputing means for participants with missing data: [A] the best mean score among the intervention arms of included trials, [B] the best mean score among the control arms of included trials, [C] the mean score from the control arm of the same trial, [D] the worst mean score among the intervention arms of included trials, [E] the worst mean score among the control arms of included trials. Using these sources of data, we developed four progressively more stringent imputation strategies. In the first example review, effect estimates were diminished and lost significance as the strategies became more stringent, suggesting the need to rate down confidence in estimates of effect for risk of bias. In the second review, effect estimates maintained statistical significance using even the most stringent strategy, suggesting missing data does not undermine confidence in the results.

**Discussion** Our approach provides rigorous yet reasonable and relatively simple, quantitative guidance that guideline developers can use for judging the impact of risk of bias as a result of missing participant data in systematic reviews of continuous outcomes.

#### 057 HANDLING TRIAL PARTICIPANTS WITH MISSING DATA IN META-ANALYSES OF DICHOTOMOUS OUTCOMES: GUIDANCE FOR SYSTEMATIC REVIEWERS

<sup>1</sup>E Akl, <sup>2,4</sup>B Johnston, <sup>3</sup>P Alonso-Coello, <sup>4</sup>I Neumann, <sup>4</sup>S Ebrahim, <sup>5</sup>M Briel, <sup>4</sup>D Cook, <sup>4</sup>G Guyatt. <sup>1</sup>American University of Beirut, Beirut, Lebanon; <sup>2</sup>Sick Kids Research Institute, Toronto, Canada; <sup>3</sup>Ibero-American Cochrane Centre, Barcelona, Spain; <sup>4</sup>McMaster University, Hamilton, Ontario; <sup>5</sup>Basel Institute for Clinical Epidemiology, Basel, Switzerland

10:1136/bmjqs-2013-002293.88

**Background** Systematic reviewers including all randomised participants in their meta-analyses need to make assumptions about the outcomes of those with missing data.

**Objectives** To provide systematic review authors with guidance on dealing with participants with missing data for dichotomous outcomes.

**Methods** The authors used an iterative process of suggesting guidance and obtaining feedback to arrive at a proposed approach.

**Results** For participants with missing data, systematic reviewers can use a range of plausible assumptions in the intervention and control arms. Extreme assumptions include 'all' or 'none' of the participants had an event, but these assumptions are not plausible. Less extreme assumptions may draw on the incidence rates within the trial (e.g., same incidence in the trial control arm) or in all trials included in the meta-analysis (e.g., highest incidence among control arms of all included trials). The primary meta-analysis may use either a complete case analysis or a plausible assumption. Sensitivity meta-analyses to test the robustness of the primary meta-analysis results should include extreme plausible assumptions. When the meta-analysis results are robust to extreme plausible assumptions, inferences are strengthened. Vulnerability to extreme plausible assumptions suggests rating down confidence in estimates of effect for risk of bias.

**Conclusions** This guide proposes an approach to establishing confidence in estimates of effect when systematic reviewers are faced with missing participant data for binary dichotomous outcomes in randomised trials.

#### 058 ASSESSMENT OF THE EVIDENCE FOR DIAGNOSTIC TESTS AND STRATEGIES: A SYSTEMATIC REVIEW OF AVAILABLE TOOLS

<sup>1,2</sup>R Mustafa, <sup>1,6,7</sup>W Wiercioch, <sup>1</sup>M Falavigna, <sup>1</sup>Y Zhang, <sup>3</sup>B Prediger, <sup>4</sup>A Cheung, <sup>1</sup>L Ivanova, <sup>5</sup>I Arevalo-Rodríguez, <sup>1</sup>H Schünemann, on behalf of the DU-Diagnosis SR group. <sup>1</sup>McMaster University, Hamilton, Canada; <sup>2</sup>University of Missouri-Kansas City, Kansas City, USA; <sup>3</sup>Germany; <sup>4</sup>University of British Columbia, Vancouver, Canada; <sup>5</sup>Universidad Nacional de Colombia, Bogota, Columbia; <sup>6</sup>Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; <sup>7</sup>National Institute of Science and Technology for Health Technology Assessment, Porto Alegre, Brazil

10:1136/bmjqs-2013-002293.89

**Background** The challenges facing guideline developers when making recommendations about diagnostic tests and strategies (DTS) are considerably different when compared to treatment recommendations.

**Objectives** To identify, describe and compare all available instruments, checklists, critical appraisal tools, and indices designed for assessing the quality of evidence (QoE) or strength of recommendations (SoR) dealing with diagnostic tests and strategies.

**Methods** We conducted a comprehensive systematic search of the literature including state of the art diagnostic guidelines, methods papers and diagnostic systematic reviews.

**Results** We identified 45 tools and modifications of existing tools to assess the QoE and SoR of DTS. Most tools acknowledge the importance of assessing the QoE and SoR separately. Most tools include individual quality criteria and study design but no tool rates all quality criteria suggested by the GRADE working group. Only two tools explicitly consider factors that increase the confidence in the evidence. When moving from evidence to recommendations, patient values and preferences and resources were rarely considered.

**Discussion** There is confusion about the terminology that describes the various factors that influence the QoE and SoR. The criteria for evaluating the QoE and moving from evidence to recommendations are incomplete for most guideline development frameworks that we evaluated.

**Implications for Guideline Developers/Users** The GRADE approach is the most complete approach encompassing all factors but users will benefit from a better description of the

evidence to recommendation framework in GRADE and clarification of issues that relate to laboratory validity parameters.

059 **NONRANDOMISED STUDIES AS A SOURCE OF COMPLEMENTARY, SEQUENTIAL OR REPLACEMENT EVIDENCE FOR RANDOMISED CONTROLLED TRIALS IN SYSTEMATIC REVIEWS AND GUIDELINES**

<sup>1</sup>H Schünemann, <sup>2</sup>P Tugwell, <sup>3</sup>B Reeves, <sup>1,4</sup>E Akl, <sup>1</sup>N Santesso, <sup>1</sup>F Spencer, <sup>5</sup>B Shea, <sup>2</sup>G Wells, <sup>6</sup>M Helfand. <sup>1</sup>McMaster University, Hamilton, Canada; <sup>2</sup>University of Ottawa, Ottawa, Canada; <sup>3</sup>University of Bristol, Bristol, UK; <sup>4</sup>American University of Beirut, Beirut, Lebanon; <sup>5</sup>Vu University Medical Center, Amsterdam, Netherlands; <sup>6</sup>Oregon Health & Science University, Oregon, USA

10:1136/bmjqs-2013-002293.90

**Background** The terms applicability, generalizability, external validity, transferability generally describe one overarching theme: can available research evidence be utilised to answer the health care questions at hand, ideally supported by a judgement about the degree of confidence in this utilisation. This concept has been called directness.

**Objectives** To offer conceptual and practical guidance to those judging directness and using research evidence from non-randomised studies (NRS).

**Methods** We used a literature review and feedback from participants of a workshop funded by the Agency for Healthcare Quality and Research and the Cochrane Collaboration.

**Results** Guideline developers can use NRS as a source of complementary, sequential or replacement evidence for randomised controlled trials (RCTs) by focusing on judgements about the population, intervention, comparison and outcomes. They use NRS to complement judgements about inconsistency, the rationale and credibility of subgroup analysis, baseline risk estimates for the determination of absolute benefits and downsides, and the directness of surrogate outcomes. Authors use NRS as sequential evidence to provide evidence when the evidence from RCTs is insufficient (e.g. long-term harms). Use of evidence from NRS may also replace RCT evidence when RCTs provide indirect evidence but NRS provide overall higher quality, direct evidence. We developed a simple tool and algorithm to make judgements about indirectness more transparent.

**Discussions** These judgements need to be made in the context of other quality of evidence domains.

**Implications for Guideline Developers/Users** The transparency of the framework will support interaction with those making health care decision and policy.

060 **APPRAISING IMPLEMENTABILITY DURING THE DEVELOPMENT PROCESS RESULTED IN GUIDELINE REVISION**

<sup>1,2</sup>S van der Veer, <sup>1</sup>A Simon, <sup>3</sup>D Weismann, <sup>4</sup>G Lehner, <sup>5</sup>L Coentrão, <sup>1</sup>D Dongelmans, <sup>6</sup>S van Laecke, <sup>6</sup>B Lapauw, <sup>7</sup>A Ortiz, <sup>8</sup>A Schiller, <sup>9</sup>V Tesar, <sup>2,6</sup>E Nagler. <sup>1</sup>Academic Medical Center, Amsterdam, The Netherlands; <sup>2</sup>European Renal Best Practice, Ghent, Belgium; <sup>3</sup>University Hospital Würzburg, Würzburg, Germany; <sup>4</sup>Medical University Innsbruck, Innsbruck, Austria; <sup>5</sup>University of Porto Sao Joao Hospital Centre, Porto, Portugal; <sup>6</sup>University Hospital Ghent, Ghent, Belgium; <sup>7</sup>IIS-Fundacion Jimenez Diaz, Madrid, Spain; <sup>8</sup>Emergency Hospital Timisoara, Timisoara Romania General University Hospital, Prague, Czech Republic

10:1136/bmjqs-2013-002293.91

**Background** The GuideLine Implementability Appraisal (GLIA) instrument has been suggested for identifying potentially remediable implementability issues during the guideline development process.

**Objective** To explore to what extent using GLIA during the development process would result in guideline revision before publication.

**Methods** The development process of the European hyponatremia guideline -coordinated by European Renal Best Practice - was our study context. Using the GLIA web-tool, eleven clinicians and methodologists from eight countries individually appraised 27 guideline statements. In a face-to-face consensus meeting, four GLIA panelists and one guideline development group (GDG) representative summarised potential implementability issues. The GDG discussed these issues, and revised the guideline if deemed necessary.

**Results** We identified 33 issues; the GDG accepted 26 as potentially hampering implementability. This resulted in statement reformulation with (n=5) and without (n=10) influencing clinical content, adding or (re)moving entire statements (n=8), and adding information to tables or rationales (n=3). The majority of issues declined by the GDG (n=7) addressed clinical situations that were covered elsewhere in the guideline or were considered to be uncommon.

**Discussion** Using GLIA during the development process resulted in a revised guideline. We felt that GDG representation in the consensus meeting optimize our appraisal process.

**Implications for Guideline Developers/Users** Guideline organizations may want to consider incorporating GLIA into their development process. This may raise GDGs' awareness of potential implementability issues, and allow revision of the guideline accordingly prior to publication. Future research should explore the effect of GLIA-based revisions on implementability as assessed by guideline users.

061 **GUIDELINE IMPLEMENTABILITY APPRAISAL (GLIA) IN US NATIONAL GUIDELINES**

<sup>1</sup>W Chan, <sup>2</sup>T Pearson. <sup>1</sup>Kaiser Permanente, Northwest, Portland, USA; <sup>2</sup>University of Rochester, Rochester, NY, USA

10:1136/bmjqs-2013-002293.92

**Background** Guidelines must be implemented in order to impact health outcomes. Identifying and addressing potential barriers to implementation during guideline development can improve implementability.

**Objectives** To describe the processes and results of embedding guideline implementability appraisal, into prominent US cardiovascular disease risk reduction guidelines.

**Methods** The GuideLine Implementability Appraisal (GLIA) tool (Yale Center for Medical Informatics), was integrated into the guideline development processes of a US national-level organisation. A member of the Implementation Science Work Group (ISWG) with prior experience in GLIA appraisals trained the Guideline Development Teams (GDTs), early in the guideline development process, with the intent of raising awareness of potential barriers to implementation so they might be addressed during guideline development. Formal GLIA appraisals were performed on the drafts of the guideline reports, by members of the ISWG, as well as volunteers from outside the guideline programme. To minimise interference with timelines, appraisals were carried out and written reports returned to the GDTs within 2 weeks of release of the draft reports.

**Results** A number of potential barriers to implementation were identified in the draft reports, such as: use of implicit terms in recommendation language, inconsistency of thresholds and terms used within a guideline, unclear applicability of assessment