Mignini et al. Reproductive Health (2016) 13:21
DOI 10.1186/s12978-016-0130-3

Reproductive Health

RESEARCH Open Access

Graphical displays for effective reporting of evidence quality tables in research syntheses

Luciano Mignini[1], Rita Champaneria[2], Ekaterina Mishanina[3], Khalid S. Khan[4*] and with the EBM-CONNECT Collaboration

Abstract

Background: When generating guidelines, quality of the evidence is tabulated to capture its several domains, often using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach. We developed a graphic display to capture deficiencies, outliers and similarities across comparisons contained in GRADE tables.

Methods: Based on a systematic literature review capturing the effects of 32 different therapeutic comparisons on dysmenorrhoea, we synthesised evidence quality in tables and graphs. We evaluated time taken to accurately assess evident quality and preference for tables vs graphs.

Results: The plots provided visually striking displays of strengths and weaknesses of the evidence across the spectrum of comparisons on a single page. Equivalent tabulated information spread over 4 pages. Participants preferred and interpreted graphs quicker and more accurately than tables.

Conclusions: The graphic approach we developed makes interpreting evidence easier. Large tables are dry and cumbersome to read and assimilate. When guideline statements are accompanied by these plots, they have the scope for improving the credibility of the recommendations made, as the strength of the evidence used can be clearly seen. Further empirical research will establish the place for graphic displays.

Keywords: Grade, Grade plots, Guidelines, Radar charts, Systematic reviews

Précis: GRADE plots allow easier, Quicker and more accurate identification of deficiencies in the quality of studies, Compared to tabulated results

Resumen: Para generar guías de práctica clínica, la calidad de la evidencia se tabula, a menudo utilizando el GRADE (Clasificación de las recomendaciones de la evaluación y desarrollo). Hemos desarrollado una pantalla gráfica para capturar deficiencias, valores atípicos y similitudes a través de comparaciones que figuran en las tablas GRADE.

Métodos: Basado en una revisión sistemática de la literatura para analizar los efectos terapéuticos de 32 diferentes tratamientos en la dismenorrea, se sintetizó calidad de la evidencia en tablas y gráficos. Evaluamos tiempo necesario para evaluar con precisión evidente calidad y preferencia por las tablas vs gráficos.

(Continued on next page)

* Correspondence: k.s.khan@qmul.ac.uk
[4]Centre for Health Sciences, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK
Full list of author information is available at the end of the article

(Continued from previous page)

Resultados: Los gráficos muestran fácilmente las fortalezas y debilidades de la evidencia en todo el espectro de las comparaciones en una sola página. Los participantes prefieren los gráficos y estos son interpretados más rápido y con más precisión que las tablas.

Conclusiones: El enfoque gráfico que hemos desarrollado hace más fácil la interpretación de la evidencia. Las tablas son difíciles y engorrosas para leer y asimilar. Cuando los estados únicamente se acompañan de estos gráficos, tienen la posibilidad de mejorar la credibilidad de las recomendaciones, ya que la fuerza de la evidencia utilizada se puede ver claramente. Más investigación empírica establecerá el lugar para pantallas gráficas.

Background

When using scientific evidence for clinical decision making, it is essential to know its quality, in order to be confident in the recommendations [18]. The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) is a quality assessment tool used to evaluate limitations of the evidence and to provide an underpinning strength to recommendations [1]. GRADE tables describe various quality parameters including study design, risk of bias, inconsistencies, indirectness and imprecision, to generate an overall rating of the evidence (from high to very low). To be comprehensive, the quality parameters (criteria, factors or domains) have to be included for each outcome and comparison separately, making the tables bulky and difficult to use [3, 20]. These could be presented as graphic displays compressing a large amount of data into concise, easy to interpret figures [12]. This article explores how addition of graphic displays of evidence quality assessment to GRADE may help readers, providing the findings of a user evaluation.

Methods

Tabulating evidence profiles

Primary dysmenorrhoea, a common idiopathic chronic pelvic pain syndrome of unknown aetiology [13], was chosen as an example. We used this topic to demonstrate the difficulties encountered when bringing together complex data on quality of the evidence on numerous comparisons. In this systematic review we searched electronic literature databases, including Medline, Embase and the Cochrane Library until January 2010 [14]. Harms alerts from relevant organizations such as the US Food and Drug Administration (FDA) and the UK Medicines and Healthcare products Regulatory Agency (MHRA) were also searched. We selected randomised controlled trials (RCTs) which were at least single blinded, with at least 80 % follow up at primary end point, and had a sample size of at least 10 women in each group.

There were many interventions compared for effect on various outcome measures. For each comparison and outcome pair, evidence was initially graded by the study design. We assigned all evidence a high level of quality as it was based on a RCT design. If there were deficiencies in the domains risk of bias, inconsistency, indirectness and effect size or its precision, the quality level was downgraded by one level (if the deficiency was classified as serious) or by two levels (if the deficiency was classified as very serious). An example of an evidence profile is shown in Table 1. The full tabulation of the evidence profile spread over 4 pages (over 1,500 words).

Graphically displaying evidence profiles

Radar charts were used to summarise data concerning several variables in a two-dimensional graph. Each chart is made up of a number of spokes or radii, each representing a variable, arranged at equal angles. Three or more quantitative variables can be represented in this way for summarising quality parameters of clinical evidence.

A radar chart, or GRADE plot, consisting of the five most important GRADE quality parameters (study design, risk of bias, inconsistency, indirectness and imprecision of effect size) was created for each comparison and clinical outcome. The length of a spoke was proportional to the magnitude of the quality of that parameter, ranging from serious deficiency (no spoke) to no deficiency (full spoke) (Table 1) [9, 12]. If all quality parameters were of high magnitude, a GRADE plot will be of symmetrical pentagon shape. However, if one of the parameters was deficient, the shape of the pentagon will be distorted and the cross-sectional area will be smaller. GRADE plots were constructed based on a variety of treatments for pelvic pain.

Overall quality of the evidence is rated by GRADE using the following categories, high, moderate, low and very low. We utilised a traffic light colour coding scheme to represent this grading scale on GRADE plots. The following colours were allocated for each rating: green for high quality evidence, yellow for moderate, red for low/very low and white where evidence had no quality rating available. Colour coding of the data used to construct the GRADE plots and the area covered within the plot provided additional visual information regarding the quality of the evidence rating. Data for GRADE plots were double checked to avoid error.

Table 1 The features of evidence grading captured in a GRADE plot (adapted from Evid Based Med 2011;16:65-9)

Grade	Design	Risk of bias	Inconsistency	Indirectness	Effect size	Evidence Quality
	Studies are either described as randomised-control trials (RCTs) or observational.	Explains the limitations of the study based on assessment of blinding and allocation process, follow-up and withdrawals, scarcity of data, other methodological concerns e.g. incomplete reporting, subjective outcomes.	Inconsistencies due to unexplained (statistical) heterogeneity. The same weakness is only downgraded once.	Presence of indirectness in the PICO elements that affect the generalisability of participants and outcomes from each study to population of interest.	Relates to imprecision of the estimated effect based on the reported odds ratios or relative risks or mean differences for comparison. This is based on the confidence intervals, sample size and number of events.	
High	Randomised controlled trial	No problems	All/most studies show similar results with or inconsistency across studies is explained by a dose response	Population and outcomes broadly generalisable	Effect size more than 5 or less than 0.2 for all studies/meta-analyses included in comparison and significant	
Moderate			Lack of agreement between studies (e.g. statistical heterogeneity between RCTs, conflicting results)		Effect size more than 2 or less than 0.5 for all studies/meta-analyses included in comparison and significant	
Low/Very low	Controlled observational study	Problem with 2 or more elements	Serious lack of agreement between studies	Some problem with 2 or more elements	Not all effect sizes more than 2 or less than 0.5 and significant; or if effects observed not significant	
Example[a]: Thiamine vs Placebo for Pelvic Pain	Randomised trial	No limitations	Consistent	Indirect	Precise	Moderate
	Initially assigned a high strength level	→ No Change	→ No change	→ Relegation	→ No change	⌟

[a]based on evidence profile shown in Fig. 1 and BMJ 2012;344:e3011 doi:10.1136/bmj.e3011

Results

Quality of evidence on pelvic pain

Figure 1 shows the quality of the evidence for agents commonly used to treat dysmenorrhoea. When comparing single agents against placebo or no treatment, we found that the quality of the evidence ranged from moderate to low/very low. Most of the comparisons assessed had a high risk of bias and none of them directly compared interventions which we were interested in or measured outcomes important to patients. Moreover, none of comparisons had an adequately large and precise effect size.

Head to head comparisons of different active agents for the treatment of pelvic pain highlighted many gaps in the evidence base. Deficiencies in quality were seen in the majority of data plotted (Fig. 1). There was only one high quality comparison (LUNA vs. diagnostic laparoscopy) that complied with all the quality items assessed. All the remaining comparisons were of low/very low quality. Non steroidal anti inflammatory drugs (NSAIDs) [2, 7, 8, 15, 19] and hormonal regulation through oral contraceptive pills [22] were significantly more effective for pain relief than placebo. Despite poor quality these interventions remain in common use [5, 21]. Furthermore, there was severe risk of bias and all but one comparison showed an inadequate effect size.

The GRADE plots immediately captured evidence quality, for example, there was moderate quality (yellow light) evidence of the effectiveness of Thiamine and Iranian herbal medicine for treating dysmenorrhoea and for the others treatment therapies there were low to very

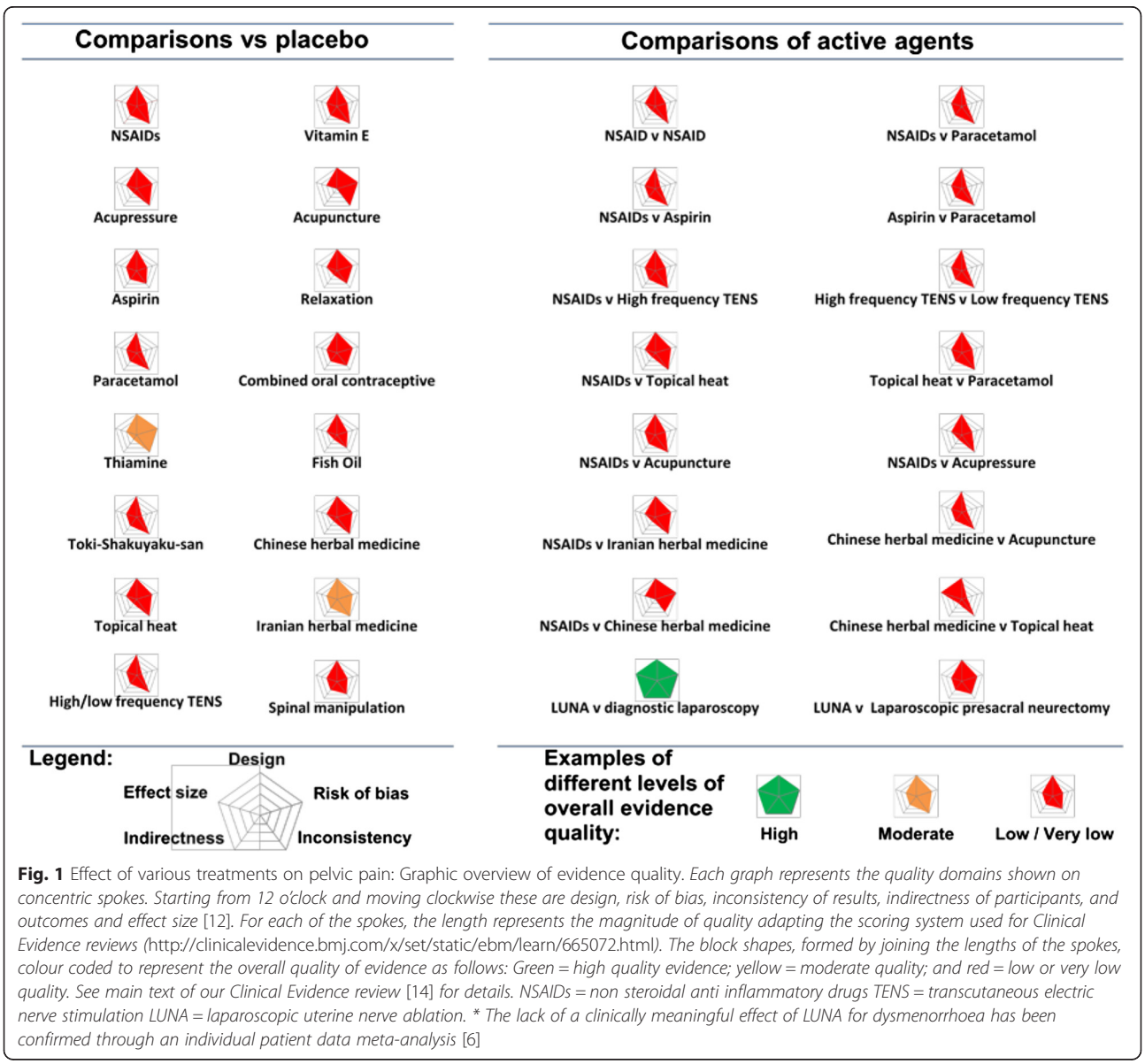

Fig. 1 Effect of various treatments on pelvic pain: Graphic overview of evidence quality. *Each graph represents the quality domains shown on concentric spokes. Starting from 12 o'clock and moving clockwise these are design, risk of bias, inconsistency of results, indirectness of participants, and outcomes and effect size* [12]. *For each of the spokes, the length represents the magnitude of quality adapting the scoring system used for Clinical Evidence reviews* (http://clinicalevidence.bmj.com/x/set/static/ebm/learn/665072.html). *The block shapes, formed by joining the lengths of the spokes, colour coded to represent the overall quality of evidence as follows: Green = high quality evidence; yellow = moderate quality; and red = low or very low quality. See main text of our Clinical Evidence review* [14] *for details. NSAIDs = non steroidal anti inflammatory drugs TENS = transcutaneous electric nerve stimulation LUNA = laparoscopic uterine nerve ablation. * The lack of a clinically meaningful effect of LUNA for dysmenorrhoea has been confirmed through an individual patient data meta-analysis* [6]

low quality evidence GRADE plots depicted in red. When therapies were compared head to head, there was only one high quality comparison (green light), which suggested that LUNA (laparoscopic uterosacral nerve ablation) should not be undertaken as a treatment for dysmenorrhoea as surgery was not found superior to no intervention. Other comparisons provided weaker recommendations, again also highlighted in red.

An empirical evaluation comparing tables vs graphs

We conducted a small randomised evaluative study to determine whether researchers and clinicians interpreted graphs quicker and more accurately than tables. Their preference for one or the other of the two approaches was also assessed. Prior to randomisation participants were shown a powerpoint presentation which explained the GRADE quality assessment tool and the new graphic concept. Seventeen participants (7 researchers and 10 hospital doctors) were then randomly assigned to either the intervention group (graphs) or the control group (tables). Participants in the intervention group were presented with a summary figure of 10 graphs summarising 5 quality parameters (study design, risk of bias, inconsistency, indirectness and imprecision). The overall quality of the evidence was indicated with the aid of colours (green for high overall quality, yellow for moderate and red for low and very low). The participants were then asked 10 questions regarding the quality of the evidence. Participants in the control group were presented with the same information and asked the same questions, but this time summarised in tables. The time taken to complete the questionnaire was recorded in seconds with the aid of a stopwatch. The preference between graphs or tables was then recorded.

We found that on average graphic displays were interpreted quicker and more accurately than tables (see Fig. 2), however the numbers of participants were too small to draw any statistically significant inferences. The majority of the participants preferred graphs to tables. Among those who received graphs 7/9 indicated a preference for graphs vs 6/8 amongst those who received tables.

Discussion

Although the GRADE quality assessment tool is not used universally, it has been utilised by some large guideline producing bodies including WHO (World Health Organization) and NICE (National Institute for Clinical Excellence). Currently, the summaries of GRADE quality assessments are presented in lengthy tables that often slow the reader down in interpretation of the findings. GRADE plots can summarise quality assessment in a more concise explicable way, making it quicker to decide on the value of the evidence. It is also possible to arrange the results of multiple interventions and outcomes in a compressed manner that can be easily examined and compared. The quality is further made explicit in the graphs by use of colour-coding, which is a strength of this approach.

There has been empirical research comparing tables and graphs of equivalent data. The compositional format and content of quantitative data displays has an impact on people's comprehension, choice and preference. [11] Our evaluative study showed that participants preferred and interpreted graphs quicker and more accurately than tables. One deficiency of our work is that we did not cover every single aspect of the GRADE approach. We also modified some aspects of GRADE to create this exemplar. For further development aspects such as publication bias and criteria for upgrading or downgrading will need to be additionally considered, while strictly adhering to the GRADE system. Another consideration should be the balance of benefit vs risk of harm. It is important to remember that the underlying concept behind the graphs is to visualise the GRADEing for the ease of assimilation by users, not to replace the in-depth analysis and consideration necessary for formulation of recommendations. Further, it is necessary to recognise the pilot or preliminary nature of our empirical evaluation. Stronger empirical work will be required to advance the advantages of graphs that show potential in our work.

Brewer et al. found that patients needed to see bar charts for a shorter amount of time compared to tables to understand the same results [4]. Bauer et al. also concluded that physicians worked significantly faster with the graphical display than tables [3]. The overall quality of the evidence can be colour coded with a traffic light system also used to display health economic data. [17]. The use of colour is not only eye-catching but if used appropriately can allow the reader to capture the overall quality immediately [11]. This idea is supported by the results of a trial by Hawley et al. who found that colour graphical representation of results (pictographs) were the most effective way of conveying information [10]. McCaffery et al. agree with these findings. Their trial reported that in adults with lower education and literacy, pictographs were the best format for displaying numerators of less than 100 (<100/1000), and bar charts were best for larger numerators (>100/1000) [16]. We therefore suggest that this strength of evidence and resulting recommendations could easily be demonstrated with a colour-coded system.

Conclusion

GRADE plots can be used to summarise large amounts of data in a concise, easy to interpret way.

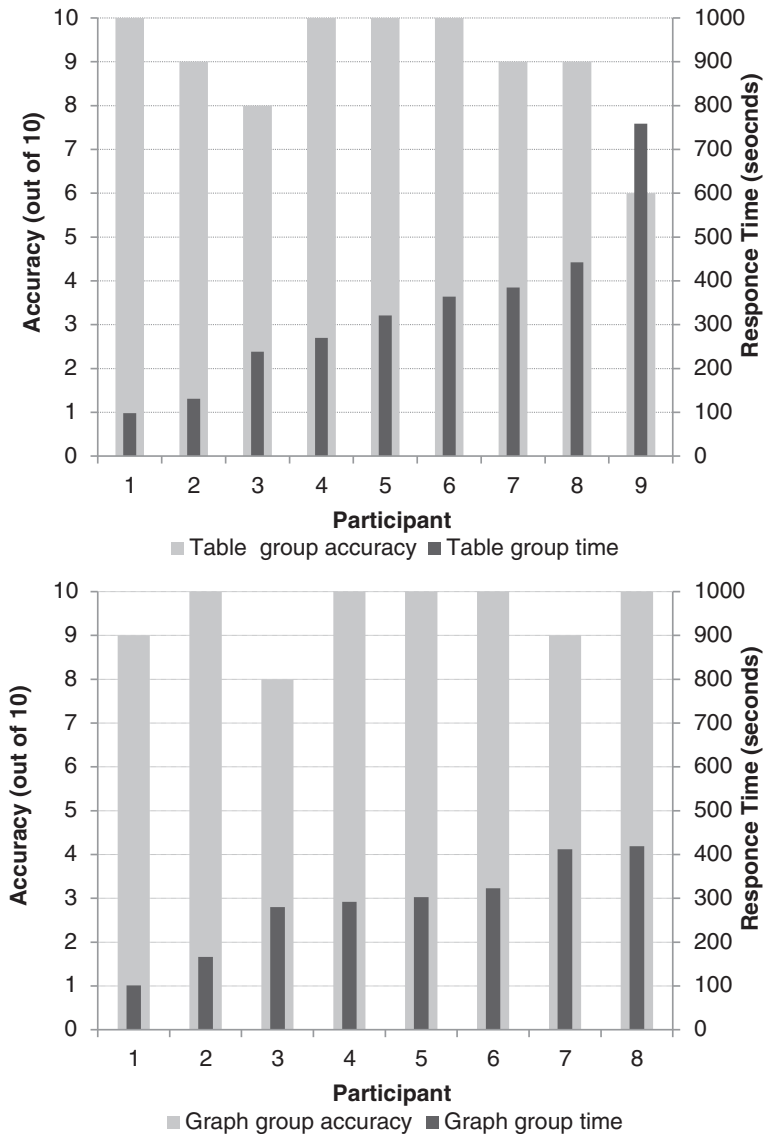

Fig. 2 Comparison of accuracy and time taken to interpret equivalent Grade tables (*above*) and graphs (*below*)

They demonstrate the quality parameters of study design, risk of bias, indirectness, inconsistencies and imprecision. The colour coded cross sectional area of the pentagon represents the overall quality of the evidence, also highlighting the strength of the recommendation. These plots provide a useful means of visually displaying evidence that could be adopted alongside the GRADE approach. The summary obtained through the plots can be read at a glance to immediately identify deficient areas that can be explored further with GRADE tables. Based on our findings, we would like to suggest to guideline makers to use graphic displays when summarising and publishing conclusions on multiple comparisons and outcomes.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
"KSK and LM conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript."

Acknowledgments
We received funding from the European Union made available to the EBM-CONNECT Collaboration through its Seventh Framework Programme, Marie Curie Actions, International Staff Exchange Scheme (Proposal no: 101377; Grant Agreement no: N° 247613); EBM-CONNECT Canadian Collaborators received funding from the Canadian Institutes of Health Research. No funders played a role in the planning and execution of this work, or in drafting of the manuscript.
The EBM-CONNECT (Evidence-Based Medicine COllaboratioN: NEtwork for systematic reviews and guideline development researCh and dissemination) Collaboration (in alphabetical order by country) includes: L Mignini, Centro

Rosarino de Estudios Perinatales, Argentina; P von Dadelszen, L Magee, D Sawchuck, University of British Columbia, Canada; E Gao, Shanghai Institute of Planned Parenthood Research, China; BW Mol, K Oude Rengerink, Academic Medical Centre, Netherlands; J Zamora, Ramon y Cajal, Spain; C Fox, J Daniels, University of Birmingham; and KS Khan, SA Tirlapur, S Thangaratinam, Barts and the London School of Medicine, Queen Mary University of London, United Kingdom.

Author details
[1]Centro Rosarino de Estudios Perinatales, Rosario, Santa Fe, Argentina. [2]Birmingham Clinical Trials Unit, School of Cancer Sciences, College of Medicine and Dentistry, University of Birmingham, Birmingham, UK. [3]Colchester University Hospital, Colchester, UK. [4]Centre for Health Sciences, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.

Received: 12 February 2016 Accepted: 19 February 2016
Published online: 09 March 2016

References

1. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S. Grading quality of evidence and strength of recommendations. BMJ. 2004;328(7454):1490.
2. Barbosa ICF. Comparative study of the efficacy and safety of valdecoxib and piroxicam in the treatment of patients with primary dysmenorrhoea. Revista Brasileira de Medicina. 2007;64:318–22.
3. Bauer DT, Guerlain S, Brown PJ. The design and evaluation of a graphical display for laboratory data. J Am Med Inform Assoc. 2010;17(4):416–24.
4. Brewer NT, Gilkey MB, Lillie SE, Hesse BW, Sheridan SL. Tables or bar graphs? Presenting test results in electronic medical records. Med Decis Mak. 2012;32:545–53.
5. Daniels JP, Khan KS. Chronic pelvic pain. BMJ. 2010;341:c4834.
6. Daniels JP, Middleton L, Xiong T, Champaneria R, Johnson NP, Lichten EM, et al. Individual patient data meta-analysis of randomized evidence to assess the effectiveness of laparoscopic uterosacral nerve ablation in chronic pelvic pain. Hum Reprod Update. 2010;16(6):568–76.
7. Daniels S, Gitton X, Zhou W, Stricker K, Barton S. Efficacy and tolerability of lumiracoxib 200 mg once daily for treatment of primary dysmenorrhoea: results from two randomised controlled trials. J Women Health. 2008;17:423–37.
8. Daniels S, Robbins J, West CR, Nemeth MA. Celecoxib in the treatment of primary dysmenorrhoea: results from two randomized, double-blind, active- and placebo-controlled, crossover studies. Clin Ther. 2009;31:1192–208.
9. Fox CE, Tirlapur SA, Gulmezoglu AM, Souza JP, Khan KS, with the EBM-Connect Collaboration. Assimilating evidence quality at a glance using graphic display: research synthesis on labor induction. Acta Obstet Gynecol Scand. 2012;91(8):885–92.
10. Hawley ST, Zikmund-Fisher B, Ubel P, Jancovic A, Lucas T, Fagerlin A. The impact of the format of graphical presentation on health-related knowledge and treatment choices. Patient Educ Couns. 2008;73:448–55.
11. Hildon Z, Allwood D, Black N. Making data more meaningful: Patients' views of the format and content of quality indicators comparing health care providers. Patient Educ Couns, 2012. Epub ahead of print.
12. Khan KS, Borowiack E, Roos C, Kowalska M, Zapalska A, Mol BW, et al. Making GRADE accessible: a proposal for graphical display of evidence quality assessments. Evid Based Med. 2011;16(3):65–9.
13. Khan KS, Champaneria R, Latthe PM. How effective are non-drug, non-surgical treatments for primary dysmenorrhoea? BMJ. 2012;344:e3011. doi:10.1136/bmj.e3011.
14. Latthe PM, Champaneria R, Khan KS. Dysmenorrhoea. Clinical Evidence. BMJ. 0813. 2011. Ref Type: Online Source
15. Marjoribanks J, Proctor M, Farquhar C, Sangkomkamhang US, Derks RS. Nonsteroidal anti-inflammatory drugs for primary dysmenorrhoea. Cochrane Database Syst Rev. 2003;4:CD001751.
16. McCaffery KJ, Dixon A, Hayen A, Jansen J, Smith S, Simpson JM. The influence of graphic display format on the interpretations of quantitative risk information among adults with lower education and literacy: a randomized experimental study. Med Decis Mak. 2012;32:532–44.
17. Nixon J, Khan KS, Kleijnen J. Summarising economic evaluations in systematic reviews: a new approach. BMJ. 2001;322(7302):1596–8. available from: PM:11431306.
18. No authors listed. The periodic health examination. Can Med Assoc J. 1979;121:(9)1193-1254.
19. Nor Azlin MI, Maryasalwati I, Norzilawati MN, Mahdy ZA, Jamil MA, Zainul Rashid MR. The efficacy of etoricoxib vs mefenamic acid in the treatment of primary dysmenorrhoea: a randomised comparative trial. J Obstet Gynaecol. 2008;28:424–6.
20. Schunemann HJ, Best D, Vist G, Oxman AD. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. CMAJ. 2003;169(7):677–80.
21. Vercellini P, Vigano P, Somigliana E. The role of the levonorgestrel-releasing intrauterine device in the management of symptomatic endometrioisis. Curr Opin Obstet Gynecol. 2005;17:359–65.
22. Wong CL, Farquhar C, Roberts H, Proctor M. Oral contraceptive pill for primary dysmenorrhoea. Cochrane Database Syst Rev. 2009;4:CD002120.