

Cow's milk allergy guidelines: a quality appraisal with the AGREE II instrument

M. Ruszczyński, A. Horvath, P. Dziechciarz and H. Szajewska

Department of Paediatrics, The Medical University of Warsaw, Warsaw, Poland

Clinical & Experimental Allergy

Summary

Background The appropriate diagnosis and management of cow's milk allergy (CMA) is challenging.

Objective To systematically review the quality of the existing guidelines on CMA.

Methods The Cochrane Library, MEDLINE, and EMBASE databases were searched from 2010 to November 2015. The methodological rigour, quality, and transparency of relevant guidelines were assessed with the use of the Appraisal of Guidelines for Research and Evaluation (AGREE II) tool.

Results Of the 15 included guidelines, two, both developed by recognized scientific organizations, achieved the highest score (100%). Eight others were considered to be of high quality (i.e., overall quality scores >60%). The quality scores for each domain varied. Of all the domains, clarity and presentation had the highest mean score, and applicability had the lowest mean score. The scores (mean $\pm$ SD) for individual domains were as follows: domain 1 (score and purpose) $62 \pm 36\%$; domain 2 (stakeholder involvement) $56 \pm 33\%$; domain 3 (rigor of development) $55 \pm 38\%$; domain 4 (clarity of presentation) $71 \pm 29\%$; domain 5 (applicability) $44 \pm 33\%$; and domain 6 (editorial independence) $60 \pm 36\%$. One guideline had the maximum possible score of 100% for all AGREE II domains.

Conclusion and Clinical Relevance A number of guidelines on CMA are available; however, their quality varies. Overall, the guidelines developed by recognized professional/scientific organizations were of the highest quality. These guidelines should be recommended for use. Still, the methodological quality of CMA guidelines may be improved.

Keywords children, cow's milk allergy, food allergy, recommendation

Submitted 23 February 2016; **revised** 20 June 2016; **accepted** 17 July 2016

Correspondence:

Marek Ruszczyński, Department of Paediatrics, The Medical University of Warsaw, Zwirki i Wigury 63A, 02-091 Warsaw, Poland.

E-mail: marek.ruszczyński@gmail.com

Cite this as: M. Ruszczyński,

A. Horvath, P. Dziechciarz and

H. Szajewska, *Clinical & Experimental Allergy*, 2016 (46) 1236–1241.

Introduction

Cow's milk allergy (CMA) is the most common food allergy in children. World-wide, the estimated prevalence is 2–5%, with most estimates of prevalence at around 3% [1–3]. Considering that the appropriate diagnosis and management of CMA may be challenging, a number of organizations have addressed that problem by publishing relevant clinical practice guidelines.

Clinical practice guidelines are 'statements that include recommendations, intended to optimize patient care, that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options' [4]. Guidelines are disseminated to promote excellence delivered by healthcare providers; therefore, their quality is critically important. To ensure

their quality, standards for the development of evidence-based guidelines have been developed [5]. Likewise, tools to assess the process of guideline development and reporting of this process in the guideline also have been developed. Our aim was to systematically review current guidelines for the management of CMA and to assess their quality, methodological rigour, and transparency using the Appraisal of Guidelines for Research and Evaluation (AGREE II) tool [5]. This review was initiated as part of the development of the Polish guidelines on oral food challenge for the diagnosis of CMA [6].

Methods

MEDLINE, EMBASE, and the Cochrane Library were searched up to May 2015, and again in November

2015, for recommendations and/or guidelines on the diagnosis and management of the CMA in children and/or adults, published in English, in the last five years. The principal search terms included ['milk' or 'milk protein' or 'dairy' or 'cow's milk' or 'cow's' in combination with 'allergy' or 'allergies' or 'intolerance' or 'hypersensitivity' or 'FPIES' or 'food protein induced proctocolitis syndrome'] and ['guidelines' or 'recommendations' or 'recommendation*' or 'guideline' or 'practice guideline*' or 'Health Planning Guideline*' or 'guidance' or 'care pathway' or 'critical pathway' or 'consensus development conference'].

The search and selection of the publications were conducted independently by three reviewers (MR, AH, and PD). All of the potentially relevant articles were retained, and their full texts were critically reviewed according to the predefined inclusion criteria.

To assess the process of guideline development and the reporting of this process in the guideline, the AGREE II instrument was used. The AGREE II instrument is a 23-item tool that addresses six guideline quality-related domains, and it is available on the website for the AGREE Research Trust (<http://www.agree-trust.org>). The first domain (scope and purpose) addresses the overall aim of the guideline, specific health questions covered by the guideline, and the guideline's target population. The second domain (stakeholder involvement) aims to answer the questions of whether the guideline was prepared by appropriate stakeholders and whether it represents the views of potential users. The third domain (rigour of development) analyses the process of collecting and synthesizing the evidence, the methodology of formulating the recommendations, and the process of their update. The fourth domain (clarity of presentation) covers the language, structure, and format of the guideline. The fifth domain (applicability) detects whether the authors identified possible barriers and facilitators to implementation as well as strategies to improve uptake and application of the guidelines. The sixth domain (editorial independence) covers conflict of interest. Each item is rated on a seven-point Likert scale ranging from 'strongly agree' to 'strongly disagree'. There are two additional assessment items (Overall Guideline Assessment) pertaining to an overall judgment of the guideline: one is, again, rated using a seven-point Likert scale and the other is a question of whether the guideline should be used, with or without modifications, or should not be used at all. The AGREE II consortium recommends that for each guideline appraisal, at least two and, preferably, four appraisers are involved.

For each domain, the scores were summed up and calculated as a percentage of the maximum possible score for that domain using the formula provided by the AGREE II consortium: [(score obtained – minimum

possible score)/(maximum possible score – minimum possible score)] × 100. The possible standardized scores range from 0% (the minimum) to 100% (the maximum). The AGREE II did not provide a minimum or maximum range for domain score quality to differentiate high- and low-quality guidelines. However, a number of previous reviewers have chosen a standardized domain score of above 60% as the threshold [7].

Three reviewers/appraisers (MR, AH, and PD) independently assessed the methodology of the guidelines using the AGREE II instrument. While none of them had previous experience with the AGREE II instrument, the reviewers underwent training from the online AGREE II tutorial and practice exercise was available at <http://www.agreetrust.org>. The AGREE II manual did not provide clarification regarding how to score for each of the seven points on the Likert scale. Therefore, the reviewers decided to discuss all scores that differed by two or more points among reviewers. These discrepancies, as well as all other disagreements between the reviewers, were resolved via discussion until a consensus was reached.

Statistical analysis

Descriptive statistics were used to describe the basic features of the data in a study. Unless otherwise specified, continuous nonparametric data were presented as a median followed by range; parametric data were presented as a mean ± standard deviation (SD). The agreement among reviewers/appraisers was calculated using the Berry–Mielke agreement coefficient *R*. Berry–Mielke is a generalization of the Cohen's kappa to an interval and ordinal measurement scales designed for more than two raters. With categorical data, *R* is equivalent to a linearly weighted kappa statistic [8]. For all analyses, we used StatsDirect statistical software (StatsDirect Ltd StatsDirect statistical software. <http://www.statsdirect.com>. England: StatsDirect Ltd 2013).

Results

For a flow diagram documenting the identification process for the eligible documents, see Fig. 1. Overall, 15 guidelines were included [2, 9–22]. Table S1 summarizes the characteristics of the excluded documents, including the reasons for exclusion.

Of the 15 included guidelines [8], were developed by international scientific organizations such as the European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) [14], World Allergy Organization (WAO) [2], and the European Academy of Allergy and Clinical Immunology (EAACI) [18]; national organizations such as the British Society for Allergy and Clinical Immunology (BSACI) [17], the American

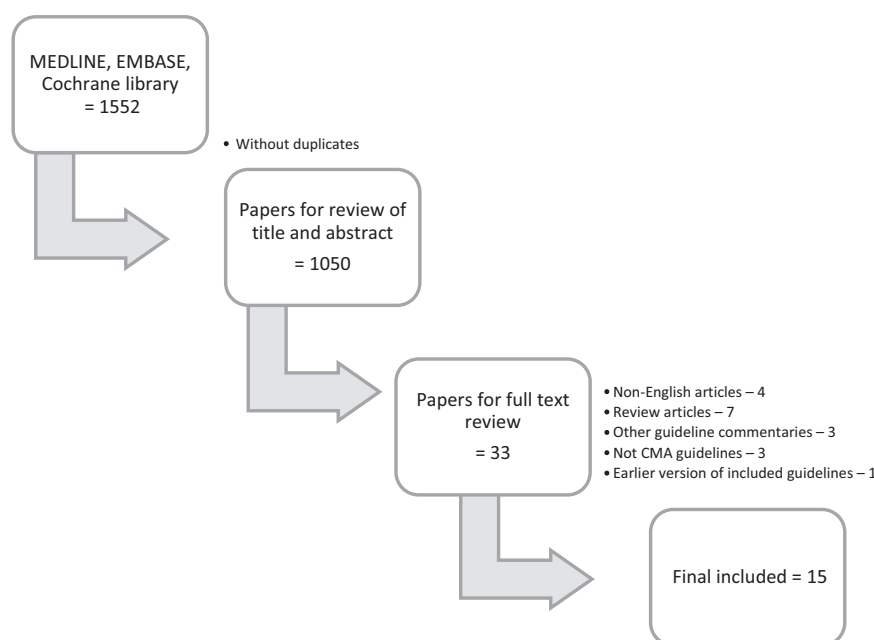


Figure 1. Flow chart (study selection).

Academy of Allergy, Asthma, and Immunology (AAAAI) [13], and the Japanese Society of Pediatric Allergy and Clinical Immunology [22]; or public bodies such as National Institute for Clinical Excellence (NICE) [12]; or through the combined efforts of an Expert Panel and Coordinating Committee representing 34 professional organizations, federal agencies, and patient advocacy groups, and the National Institute of Allergy and Infectious Diseases (NIAID) [11]. The remaining guidelines were published by expert groups [19, 20] or national health organizations and national experts [9, 10, 15, 16, 21]. Except for one guideline [16], all guidelines were developed in high-income countries. The overall agreement beyond chance among the three reviewers was 70% (Berry–Mielke agreement coefficient $R = 0.70$, which represents high agreement [$P < 0.0001$]).

The AGREE II quality scores

Domain scores. Domain scores and the overall assessment of the included CMA guidelines are summarized in Tables 1 and S2, respectively. Overall, the mean $\pm$ SD scores for various domains ranged from $43.5\% \pm 33.1\%$ to $70.8\% \pm 28.6\%$. Domain 4 (clarity and presentation) was scored the highest, and domain 5 (applicability) was scored the lowest.

Domain 1 (Scope and purpose)—The mean $\pm$ SD score for the scope and purpose domain was $62\% \pm 36\%$ (median: 76%; range: 0–100%). Two guidelines received the highest (100%) score [12, 18], and seven received

scores below 60% (ranging from 0% to 59%) [10, 13, 16, 19–22]. Such low scores were due to a lack of proper reporting.

Domain 2 (Stakeholder involvement)—For this domain, the mean $\pm$ SD score was $56\% \pm 33\%$ (median: 57%; range: 4–100%). One guideline received the highest score for this domain (100%) [18], and nine guidelines received scores below 60% (ranging from 4% to 59%) [10, 13–16, 19–22]. The main reason for the low scores was an inadequate amount of data to assess.

Domain 3 (Rigour of development)—The mean $\pm$ SD score for this domain was $55\% \pm 38\%$ (median: 50%; range: 0–100%). Two guidelines received the highest score (100%) [2, 18], and eight received scores below 60% (ranging from 0% to 50%) [10, 13–16, 20–22]. Such low scores were due to a lack of enough information provided.

Domain 4 (Clarity of presentation)—The mean $\pm$ SD score for the clarity of presentation domain was $71\% \pm 29\%$ (median 80%; range 17–100%). Three guidelines received the highest (100%) score [2, 12, 17]. Even though this domain was the most effectively addressed of all the domains, six guidelines received scores below 60% (ranging from 17% to 52%) [10, 13–16, 22]. These low scores were due to limited reporting.

Domain 5 (Applicability)—The mean $\pm$ SD score for the applicability domain was $43.5\% \pm 33\%$ (median: 32%;

Table 1. Domain scores and overall assessment of CMA guidelines using the AGREE II instrument (percentages)

Guidelines	Domain scores (%)						Overall quality	Overall assessment
	1 Scope and purpose	2 Stakeholder involvement	3 Rigour of development	4 Clarity and presentation	5 Applicability	6 Editorial independence		
BSACI (2014) [17]	96	85	99	89	67	97	100	Yes
EAACI (2014) [18]	100	100	100	100	100	100	100	Yes
WAO (DRACMA) (2010) [2]	89	96	100	100	100	100	94	Yes
NIAID-Sponsored Expert Panel (2011) [11]	89	91	88	96	31	83	89	Yes
NICE (2011) [12]	100	96	97	100	89	89	89	Yes
Singapore Ministry of Health (2010) [9]	87	72	66	94	50	0	78	YWM
Sampson <i>et al.</i> (2014) [19]	39	54	85	98	65	86	78	Yes
UK primary care practical guide (2013) [15]	87	41	50	50	24	36	72	YWM
Consensus for the Middle East (2014) [20]	22	37	42	65	19	25	66	YWM
AAAAI/EAACI PRACTALL (2012) [13]	59	59	38	44	33	58	61	YWM
ESPGHAN (2012) [14]	76	28	28	44	19	69	56	YWM
Italian guidelines (2010) [10]	56	57	19	52	3	78	50	YWM
Japanese guidelines (2014) [22]	9	11	0	33	32	0	33	No
French Society of Paediatrics (2014) [21]	0	15	10	80	15	64	22	YWM
Indonesian guidelines (2013) [16]	17	4	4	17	6	11	17	No
Mean (SD)	61.7 (35.6)	56.4 (33)	55 (37.8)	70.8 (28.6)	43.5 (33.1)	59.7 (36.3)	67 (27)	
Median	76	57	50	80	32	69	72	
Range	0–100	4–100	0–100	17–100	6–100	0–100	17–100	

CMA, cow's milk allergy; AGREE II, Assessment of Guidelines for Research and Evaluation II tool; BSACI, British Society for Allergy and Clinical Immunology; EAACI, European Academy of Allergy and Clinical Immunology; WAO, World Allergy Organization; DRACMA, Diagnosis and Rationale for Action against Cow's Milk Allergy; NIAID, National Institute of Allergy and Infectious Diseases; NICE, National Institute for Health and Care Excellence; AAAAI, American Academy of Allergy, Asthma, and Immunology; EAACI, The European Academy of Allergy and Clinical Immunology; PRACTALL, Practical Allergy; ESPGHAN, The European Society for Paediatric Gastroenterology, Hepatology and Nutrition. Yes, recommended for use; YWM, recommended for use with modifications; No, not recommended.

range: 6–100%). Two guidelines received the highest score (100%) [2, 18]. However, 10 received scores below 60% (ranging from 3% to 50%) [9–11, 13–16, 20–22]. These low scores were due to limited reporting.

Domain 6 (Editorial independence)—The mean $\pm$ SD score for this domain was 60% $\pm$ 36% (median: 69%; range: 0–100%). Two guidelines received the highest possible score (100%) [2, 18], and one guideline received the lowest possible score (0%) [22].

Overall quality scores. Table 1 also summarizes the overall quality assessment. Of 15 included guidelines, 10 (66.6%) were considered to be of high quality (i.e. scored > 60%) [2, 9, 11–13, 15, 17–20]. Only two (13.3%) guidelines, both developed by recognized professional organizations, received a score for overall

quality as high as 100% [17, 18]. Five guidelines failed to reach a score of 60% [10, 14, 16, 21, 22]. The lowest score for overall quality was 17% and was given to the national Indonesian guidelines [16].

Applicability. Six (40%) of the guidelines were categorized as 'recommended' [2, 7, 11, 12, 17–19], (47%), as 'recommended with modifications' [9, 10, 13–15, 20, 22], and two (13.3%) [16, 22], as 'not recommended'.

Discussion

Summary of findings

We aimed to systematically review the quality of the guidelines on CMA using the AGREE II instrument. Fifteen guidelines published in English met our inclusion

criteria. Of 15 included guidelines, two guidelines, both developed by the recognized scientific organizations such as BSACI and EAACI, were evaluated as being of very high quality (i.e. overall quality scores 100%). Seven others were considered to be of high quality (i.e. overall quality scores > 60%). The quality scores for each domain varied. Of all of the domains, clarity and presentation (domain 4) had the highest mean score and applicability (domain 5) had the lowest mean score. One guideline, developed by EAACI, had the maximum possible score of 100% in all AGREE II domains [18]. The WAO guidelines also achieved the maximum score for all but two domains [2]. As a result of the appraisal, six (40%) of the guidelines were categorized as 'recommended', and two were categorized as definitely 'not recommended'.

Limitations

Our search was limited to guidelines published in the last five years. While different opinions exist, most authors suggest that an update is generally required after 3–5 years [23]. Hence, guidelines published earlier would not have been up-to-date and would not be relevant to our appraisal. We limited our search to English language guidelines. Such an approach can introduce language bias. However, non-English language guidelines, while being important, may have somewhat lower of an impact on practice. The reviewers/appraisers were not blinded to the authors/organization who developed the guidelines, which may have been a potential source of bias. However, most of the guidelines were well known to the authors, so true blinding was not feasible. The reviewers/appraisers had no previous experience with the AGREE II instrument. Therefore, all reviewers undertook the recommended AGREE II consortium online training, similarly to other AGREE evaluation groups [24, 25]. Finally, while the AGREE II tool is valid and reliable, and is recommended for the appraisal of guidelines, this instrument has its limitations [24, 26]. For example, there is a lack of clear criteria for applying each point on the scale. It is also unclear how scores correspond to recommendation decisions, resulting in some subjectivity in the reviewer's interpretation of the scores.

Comparison with other studies

A number of guidelines developed by a wide variety of organizations, from different geographical regions, have been appraised using the AGREE II instrument [24–26].

Clearly, differences between the guidelines exist. However, overall, a substantial portion of the published guidelines remains of average quality. It is noteworthy that at least one previous study showed an improvement in the overall quality of published guidelines over time (in the years 2001–2006 compared with 2007–2009) [25]. This suggests a growing recognition of the principles of guideline development among guideline development groups.

Conclusions and implications

Our review shows that the quality of CMA guidelines published from 2010 through November 2015 and assessed using the AGREE II instrument varies. Only some of the published guidelines were of high quality. However, it is reassuring that these were the guidelines developed by recognized professional/scientific organizations such as BSACI, EAACI, or WAO, which have great experience in developing guidelines. These guidelines should be recommended for use in clinical practice.

Clinical practice guidelines play a major role in healthcare decision-making. Our review adds to current evidence that to ensure the quality of guidelines, standards for their development have to be followed. If the standards are strictly adhered to, unbiased and trustworthy guidelines are likely to follow. Guideline developers should adhere to standards such as those described in the AGREE II instrument to ensure the development of high-quality guidelines.

Declaration of all sources of funding

This study was funded in full by The Medical University of Warsaw.

Authors' contributions

All authors initially conceptualized this study. MR, PD, and AH were responsible for data collection, data analysis, data interpretation, and preparation of the report. MR and HS assumed the main responsibility for the writing of this manuscript. All authors contributed to (and agreed upon) the final version.

Conflict of interest

The authors declare no conflict of interest.

References

- Nwaru BI, Hickstein L, Panesar SS *et al.* Prevalence of common food allergies in Europe: a systematic review and meta-analysis. *Allergy* 2014; **69**:992–1007.
- Fiocchi A, Brozek J, Schünemann H *et al.* World Allergy Organization (WAO) diagnosis and rationale for action against cow's milk allergy (DRACMA) guidelines. *Pediatr Allergy Immunol* 2010; **21**:1–125.
- Rona RJ, Keil T, Summers C *et al.* The prevalence of food allergy: a meta-analysis. *J Allergy Clin Immunol* 2007; **120**:638–46.
- Graham R, Mancher M, Wolman DM, Greenfield S, Steniberg E. *Clinical practice guidelines we can trust*. Washington, DC: National Academies Press, 2011.
- AGREE Next Steps Consortium (2009). *The AGREE II Instrument* [Electronic version]. Available from <http://www.agreetrust.org>. (Last accessed 1 January 2016).
- Jarocka-Cyrta E, Nowak-Węgrzyn A, Ruszczyński M *et al.* Oral food challenge for diagnosis of cow's milk allergy. Statement of the Food Allergy Working Group of the Polish Society for Paediatric Gastroenterology, Hepatology and Nutrition. *Standards Med Pediatr* 2015; **12**:501–16.
- Sanclemente G, Acosta JL, Tamayo ME, Bonfill X, Alonso-Coello P. Clinical practice guidelines for treatment of acne vulgaris: a critical appraisal using the AGREE II instrument. *Arch Dermatol Res* 2014; **306**:269–77.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**:159–74.
- Lee BW, Aw MM, Chiang WC *et al.* Academy of medicine, Singapore-Ministry of Health clinical practice guidelines: management of food allergy. *Singapore Med J* 2010; **51**:599–607.
- Caffarelli C, Baldi F, Bendandi B *et al.* EWGPAG. Cow's milk protein allergy in children: a practical guide. *Ital J Pediatr* 2010; **15**:5.
- NIAID-Sponsored Expert Panel, Boyce JA, Assa'ad A *et al.* Guidelines for the diagnosis and management of food allergy in the United States: report of the NIAID-sponsored expert panel. *J Allergy Clin Immunol* 2010; **126**:S1–58.
- Centre for Clinical Practice at NICE (UK). Food Allergy in Children and Young People: Diagnosis and Assessment of Food Allergy in Children and Young People in Primary Care and Community Settings. London: National Institute for Health and Clinical Excellence (UK); 2011 (NICE Clinical Guidelines, No. 116) Available from <http://www.ncbi.nlm.nih.gov/books/NBK82184/>.
- Sampson HA, Gerth van Wijk R, Bindslev-Jensen C *et al.* Standardizing double-blind, placebo-controlled oral food challenges: American Academy of Allergy, Asthma & Immunology-European Academy of Allergy and Clinical Immunology PRACTALL consensus report. *J Allergy Clin Immunol* 2012; **130**:1260–74.
- Koletzko S, Niggemann B, Arato A *et al.* European Society of Pediatric Gastroenterology, Hepatology, and Nutrition. Diagnostic approach and management of cow's-milk protein allergy in infants and children: ESPGHAN GI Committee practical guidelines. *J Pediatr Gastroenterol Nutr* 2012; **55**:221–9.
- Venter C, Brown T, Shah N, Walsh J, Fox AT. Diagnosis and management of non-IgE-mediated cow's milk allergy in infancy – a UK primary care practical guide. *Clin Transl Allergy* 2013; **3**:23.
- Munasir Z, Muktiarti D. The management of food allergy in Indonesia. *Asia Pac Allergy* 2013; **3**:23–8.
- Luyt D, Ball H, Makwana N *et al.* Standards of Care Committee (SOCC) of the British Society for Allergy and Clinical Immunology (BSACI). BSACI guideline for the diagnosis and management of cow's milk allergy. *Clin Exp Allergy* 2014; **44**:642–72.
- Muraro A, Werfel T, Hoffmann-Sommergruber K *et al.* EAACI food allergy and anaphylaxis guidelines: diagnosis and management of food allergy. *Allergy* 2014; **69**:1008–25.
- Sampson HA, Aceves S, Bock SA *et al.* Food allergy: a practice parameter update-2014. *J Allergy Clin Immunol* 2014; **134**:1016–25.
- Vandenplas Y, Abuabat A, Al-Hammadi S *et al.* Middle east consensus statement on the prevention, diagnosis, and management of cow's milk protein allergy. *Pediatr Gastroenterol Hepatol Nutr* 2014; **17**:61–73.
- Dupont C, Chouraqui JP, de Boissieu D *et al.* Dietary treatment of cows' milk protein allergy in childhood: a commentary by the Committee on Nutrition of the French Society of Paediatrics. *Br J Nutr* 2012; **107**:325–38.
- Urisu A, Ebisawa M, Ito K *et al.* Japanese guideline for food allergy 2014. *Allergol Int* 2014; **63**:399–419.
- Vernooij RW, Sanabria AJ, Solà I, Alonso-Coello P, Martínez García L. Guidance for updating clinical practice guidelines: a systematic review of methodological handbooks. *Implement Sci* 2014; **9**:3.
- Lee GY, Yamada J, Kyololo O, Shorkey A, Stevens B. Pediatric clinical practice guidelines for acute procedural pain: a systematic review. *Pediatrics* 2014; **133**:500–15.
- Acuña-Izcaray A, Sánchez-Angarita E, Plaza V *et al.* Quality assessment of asthma clinical practice guidelines: a systematic appraisal. *Chest* 2013; **144**:390–7.
- Polus S, Lerberg P, Vogel J *et al.* Appraisal of WHO guidelines in maternal health using the AGREE II assessment tool. Garner P, ed. *PLoS ONE* 2012; **7**:e38891.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Characteristics of excluded studies and reasons for exclusion.

Table S2. Assessment of CMA guidelines using the AGREE II instrument – all questions (percentages).