BACKGROUND

Cow’s milk protein allergy (CMPA) is an immune-mediated allergic response to proteins in milk, one of the most common infant food allergies with a reported incidence by age 2 in the UK of 2.16%–2.4%.\(^1,2\) CMPA is classified as immunoglobulin E (IgE)-mediated, non-IgE-mediated or mixed according to the underlying immunological mechanisms. IgE-mediated reactions typically present immediately after ingestion, with pruritus, angioedema and anaphylaxis.\(^3,4\) Seventy percent of cases are classified as non-IgE-mediated, which is more

Symptom scores in the diagnosis of pediatric cow’s milk protein allergy: A systematic review

Georgina Thompson\(^1\) | Zhivko Zhelev\(^1\) | Jaime Peters\(^1\) | Sara Khalid\(^2\) | Simon Briscoe\(^2\) | Liz Shaw\(^2\) | Michael Nunns\(^2\) | Sian Ludman\(^3\) | Christopher Hyde\(^1\)

Abstract

Background: Cow’s milk protein allergy (CMPA) is an immune-mediated allergic response to proteins in milk that is common in infants. Broad CMPA symptoms make diagnosis a challenge, particularly in primary care. Symptom scores may improve a clinician’s awareness of symptoms, indicating a need for further testing. This systematic review examined the development and evaluation of such symptom scores for use in infants.

Methods: CENTRAL, MEDLINE, EMBASE and CINAHL databases were searched from inception to 3 December 2019 (Updated 14 November 2020) for diagnostic accuracy studies, randomised controlled trials, observational studies, economic evaluations, qualitative studies and studies reporting development of the tools. Data were not suitable for meta-analysis due to clinical and methodological heterogeneity, so were narratively synthesised.

Results: We found two symptom scores evaluated in one and fourteen studies, respectively. Estimated sensitivity and specificity ranged from 37% to 98% and 38% to 93%. The evaluations of each tool were at high risk of bias or failed to address issues such as clinical and cost-effectiveness.

Conclusions: Estimates of accuracy of symptom scores for CMPA offered so far should be interpreted cautiously. Rigorous, conflict-free research based on well-defined roles for the tools is urgently required.

KEYWORDS

cow’s milk allergy, symptom score, systematic review
commonly associated with delayed, generalised symptoms including faltering growth and gastrointestinal and dermatological manifestations, which can take up to 48 hours to develop.\(^1\) Subsets of non-IgE-mediated CMPA include eosinophilic oesophagitis, food protein-induced enterocolitis and allergic proctocolitis, which are less common and present with a constellation of gastrointestinal symptoms.\(^5\)

The double-blind, placebo-controlled food challenge (DBPCFC) is considered the gold standard for diagnosis of food allergy, but it is time-consuming and may not be readily available in many clinical settings.\(^4,6\) Whereas IgE-mediated and mixed reactions may be identified by skin prick testing (SPT), food allergen-specific serum IgE (sIgE) and oral food challenges (OFC), non-IgE-mediated reactions are more difficult to identify with currently available techniques.\(^3,4,7,8\)

National Institute for Clinical Excellence (NICE) and Diagnosis and Rationale for Action against Cow’s Milk Allergy (DRACMA) guidelines recommend that where CMPA is suspected, clinicians consider taking an allergy-focused history before deciding whether an exclusion/inclusion trial is warranted.\(^2,5\) Delays in appropriate diagnosis of CMPA reflect both poor awareness of the condition and insufficient capacity in primary care to effectively apply the recommended diagnostic strategy. If exclusion diets are trialled, patients are often left on unnecessary exclusions without the recommended reintroduction of cow’s milk to prove CMPA.\(^7,8\)

Accurate and timely identification of CMPA will improve outcomes for patients.\(^6,10\) Symptom scores, such as the commercially available Cow’s Milk-related Symptom Score (CoMiSS\(^{TM}\)), have been created to be used in the clinical setting where an abundance of specific symptoms is interpreted as a possible indicator of CMPA.\(^11\) Despite being widely available online, there is limited information on the development, validity and accuracy of symptom scores for CMPA. Evaluating their performance will help clinicians to improve the early stages of the process of CMPA diagnosis. We have systematically reviewed the literature on the development and evaluation of symptom scores used in the diagnostic work-up of CMPA.

## METHODS

We followed Cochrane Collaboration recommendations for best practice,\(^12\) and the review protocol is registered on PROSPERO (registration number CRD42020165606).

### 2.1 Inclusion and exclusion criteria

We included diagnostic accuracy studies, randomised controlled trials and observational studies, economic evaluations, qualitative studies and systematic reviews, reporting on the development and evaluation of symptom score tools for suspected CMPA. Study subjects included infants aged 0–36 months. Case reports, conference abstracts and meeting reports were excluded.

### 2.2 Search strategy

The search strategy was developed by GT and SB (information specialist) and executed by GT. CENTRAL, MEDLINE, EMBASE and CINAHL were searched from inception to 3 December 2019 without language or study type restrictions (Updated 14 November 2020). We combined indexing and title and abstract keyword terms for ‘milk’, ‘al- lergy’, ‘infant’ and ‘score’ (see Appendix S1). We searched the World Health Organization International Clinical Trials Registry Platform. Forward citation searches on included studies were conducted using Science Citation Index, and reference lists were manually inspected.

### 2.3 Study selection

Two authors (GT, SK) independently reviewed the titles and abstracts, and subsequently, the full texts of studies were identified from the literature search. Interpretative discrepancies were resolved by discussion. Study authors were contacted as needed for clarification (see Appendix S2).

### 2.4 Data extraction and quality appraisal

Standardised data extraction form was developed and piloted (GT, ZZ) (Appendix S3). GT extracted all data, which were then checked by a second reviewer (SK). Three authors (GT, ZZ and JP) independently assessed the methodological quality of the included studies and resolved discrepancies through discussion. We used QUADAS-2 for test accuracy studies and PROBAST for studies reporting the development and piloting of the symptom scores, as it was felt this tool would be the most useful tool to appraise their quality.\(^13,14\) Studies in which a symptom score was used to assess the effectiveness of specialist milk formula were included for completeness, but since they did not report on the performance of the tool, no quality assessment was carried out.

### 2.5 Data analysis, synthesis and reporting

Symptom scores were analysed separately and compared qualitatively. Results from test accuracy studies were summarised by

---

**Key Message**

Symptom scores have been created to help clinicians to recognise the symptoms of cow’s milk protein allergy in infants, thus aiding earlier diagnosis. This systematic review found that there is no well-defined role for these tools and that current estimates of their accuracy should be interpreted with caution.
the number of false positive, false negative, true positive and true negative results, and sensitivity and specificity estimates. Forest plots were generated for each test pair of sensitivity and specificity across studies using Review Manager 5.4 (The Cochrane Collaboration 2020). Data were not pooled due to the small number of studies reporting both sensitivity and specificity, and extensive clinical and methodological heterogeneity among included studies.

3 | RESULTS

3.1 | Study selection

A total of 1139 studies were identified (excluding duplicates) with 84 eligible for full-text review. Fourteen studies met our inclusion criteria. One further study was identified through citation searches and one through the updated search. A total of 16 studies were included in this review. The study selection process is shown in Figure 1.

3.2 | Included studies

Two symptom scores were evaluated in the 16 included studies (see Table 1). One test accuracy study from the United States evaluated a CMPA questionnaire. The remaining fifteen studies focused on CoMiSS™ and were conducted in Belgium (n = 7), Italy (n = 2), Turkey (n = 2) and one in each of Poland, United Kingdom, India and China; these consisted of four studies on aspects of the development and piloting of CoMiSS™, six test accuracy studies and five effectiveness studies using CoMiSS™ as an outcome measure.

3.3 | Description of included symptom scores

3.3.1 | CoMiSS™

Cow’s Milk-related Symptom Score (CoMiSS™), previously symptom-based score, was developed by consensus of 18 experts from 14 different hospital sites in Belgium to help diagnose mainly non-IgE-mediated CMPA. It includes gastrointestinal symptoms (regurgitation, altered stool composition), skin manifestations (eczema, urticaria), respiratory tract symptoms and general symptoms such as crying time. The overall score ranges from 0 to 33, with each symptom having a maximum score of six, apart from respiratory symptoms, with a maximum score of three. An arbitrary cut-off point of >12 was originally selected as the criterion to highlight infants at risk of CMPA who require further testing by exclusion diet and reintroduction, a score of which would require the presence of at least two severe symptoms. Full details of CoMiSS™ can be found in Appendix S4. CoMiSS™ is made available online through Nestlé Health Science.

3.3.2 | CMPA questionnaire

Gibbons et al.16 aimed to evaluate a multisystem questionnaire that would help diagnose non-IgE-mediated CMPA and be easy to apply.
and score in a clinical setting. Symptoms for the questionnaire were selected based on chart reviews of patients diagnosed with CMPA and a literature review of its clinical manifestations in infants. Each symptom scores 1 for a positive response and 0 for a negative response, which contribute to an overall score. Vomiting was scored based on frequency. Receiver operating curve (ROC) analysis was used to determine a cut-off point that results in balanced performance in terms of sensitivity and specificity. Full details of the questionnaire can be found in Appendix S5.

### Methodological quality of included studies

#### 3.4.1 Development and piloting studies

Risk of bias was high as the studies recruited presumed healthy infants whose CMPA status was deemed to be negative without confirmatory tests, and no clear test or criteria for exclusion of those not having CMPA were reported.18–20 These studies were also poorly...
reported making them difficult to assess risk of bias; for example, there was a lack of clarity on who completed the CoMiSS™ and whether they were blinded to allergy status of the infant, and no details were provided for infants who scored >12.

3.4.2 | Diagnostic test accuracy studies

The quality of included diagnostic accuracy studies is summarised in Table 3. All studies scored high risk in at least two domains. Four studies scored high risk in the patient selection domain: two used a case-control study design,16,21 and two selected only infants that scored a CoMiSS™ of >12.22,23 Four studies scored high risk for index test interpretation bias, due to the lack of pre-specified threshold criterion for a positive screen and a lack of blinding to the results of the reference standard.16,22–24 All studies scored high risk of bias for not using DBPCFC as the reference standard.16,21-26 All studies scored high risk of bias for flow and timing: one left an interval of up to 3 months between applying CoMiSS™ and conducting OFC,25 three inappropriately excluded infants from the analysis,22 and three used different reference standards to establish CMPA.21,22,26 Concerns regarding applicability of the studies included selecting infants with a prior diagnosis of CMPA,21 using the tool to screen unselected infants without prior suspicion of CMPA22 and performing the index test after an elimination diet rather than at presentation.21,23 Full details on the methodological quality of these studies can be found in Appendix S6.

3.4.3 | Funding

Nine studies acknowledged conflicts of interest with regard to authors’ connections to pharmaceutical or milk formula industries, including companies that have been involved in the development of CoMiSS™.11,19,20,23,26-30 Six studies received direct funding from the pharmaceutical or milk formula industry.11,23,27-30

3.5 | Main findings

3.5.1 | CoMiSS™

**Development and piloting studies**

Four studies reported on the development or initial validation of the CoMiSS™ tool.11,18-20 The development study stated that the predictive value of CoMiSS™ was 80% if the score was >12 at the start and decreased to <6 under an elimination diet with extensive hydrolysate formula.11 Consensus by an expert panel was reached for all five SBS items, with a view that CoMiSS™ could be used as an ‘awareness’ tool for CMPA.

Two studies investigated the performance of CoMiSS™ in healthy infants aged <6 months to provide a scientific basis for the recommended CoMiSS™ cut-off of >12. Vandenplas reported a median CoMiSS™ score of 3 in a cohort of 413 infants from Belgium (31.2%), Italy (18.2%), Poland (19.1%) and Spain (31.5%); median crying (p < .001), regurgitation (p = .001) and eczema (p = .039) scores differed significantly across the age categories.20 Bigorajska reported a median CoMiSS™ score of 4 in a cohort of 226 infants in Poland; similarly, age impacted on individual crying (p = .001) and stool (p < .001) scores.18

Vandenplas investigated the inter-rater (healthcare professional (HCP) vs. parent) variability of CoMiSS™ in 148 Spanish infants.19 Excellent agreement with an intraclass correlation coefficient (ICC) .981 was reported (95% CI: 0.974–0.986, p < .001). In the second phase of the study, day-to-day variability in 72 Belgian infants (parents assessment over 3 consecutive days vs HCP) was excellent for parental prospective scores, ICC .93 (95% CI: 0.90–0.96; p < .001), but poorer between the HCP and parents on Day 1 versus Day 2, ICC .53 (95% CI: 0.34–0.68; p < .001). The authors suggest that CoMiSS™ can reliably be scored by parents without additional training.

**Test accuracy studies**

Test accuracy of CoMiSS™ was evaluated in 6 studies,16,21,23–26 Study characteristics and accuracy results are summarised in Tables 2 and 3, respectively. Additional accuracy data are reported in Appendix S7. Forest plots are shown in Appendix S8.

Two studies evaluated the accuracy of CoMiSS™ to predict a positive food challenge following an elimination diet. Sirin Kos examined CoMiSS™ in infants aged 0–10 months with a prior diagnosis of CMPA against OFC or skin prick test in infants ≤6 months of age and reported that a ≥50% decrease in CoMiSS™ score had sensitivity of 84% (95% CI: 70% to 93%).21 However, the lack of a healthy infant arm in this two-gate study design leads to potentially misleading accuracy estimates since high sensitivity naturally leads to low specificity and so would generate more false-positive results. Vandenplas evaluated CoMiSS™ in 85 formula-fed infants aged 0–6 months with a baseline CoMiSS™ score of ≥12 (mean ± SD 13.65 ± 1.75; range 12–21) and reported that a decrease in CoMiSS™ score to <6 after 1-month elimination diet had sensitivity of 76% (95% CI: 63% to 86%) and specificity of 58% (95% CI: 37% to 77%) in predicting a positive OFC.23

Four studies evaluated the test accuracy of CoMiSS™ at presentation relative to results of confirmatory tests for CMPA. Prasad evaluated CoMiSS™ against either OFC or an immunology test (ImmunoCAP) in 38 infants aged 0–24 months and reported moderate sensitivity of 79% (95% CI: 67% to 87%) but low specificity of 38% (95% CI: 14% to 68%).26 Zeng examined CoMiSS™ against OFC in infants aged 1–12 months and through ROC analysis reported moderate-to-high sensitivity and specificity of 88% (95% CI: 68% to 97%) and 79% (95% CI: 49% to 95%), respectively, with a best diagnostic cut-off point of 5.5.24 Salvatore examined CoMiSS™ against response to cow’s milk-free diet without performing an OFC, in 47 infants aged 1–12 months. A cut-off score of 12 yielded poor sensitivity of 37% (95% CI: 16% to 62%) but high specificity of 93% (95% CI: 75% to 99%).25 Improved sensitivity of 84% (95% CI: 60% to 97%) and only slightly reduced specificity of 85% (95% CI: 67% to 96%) were
| Study          | Inclusion criteria and selection                                                                 | Index test cut-off       | Index test examiner     | Reference standard(s)       | Definition of CMPA by reference standard                                                                 |
|---------------|-------------------------------------------------------------------------------------------------|--------------------------|-------------------------|-----------------------------|----------------------------------------------------------------------------------------------------------|
| CMPA questionnaire |                                                                                                 |                          |                         |                             |                                                                                                          |
| Gibbons, 2012 | Case-control: infants <2 years seen at gastroenterology clinic with a diagnosis of CMPA & presumed healthy controls seen at routine health visit | ROC-determined: score ≥6 | Health professional     | CMFD + ImmunoCAP             | Positive response to CMFD + test negative for IgE-mediated disease + failed management of alternative conditions |
| CoMiSS™      |                                                                                                 |                          |                         |                             |                                                                                                          |
| Prasad, 2018 | Infants aged 0–24 months seen at paediatric clinic with one or more symptoms of CMPA, including cutaneous, respiratory or GI | Pre-specified: score ≥12 | General paediatrician   | OFC or ImmunoCAP             | Positive if symptoms reappear within 2 weeks of OFC or IgE-mediated CMPA confirmed by ImmunoCAP           |
| Salvatore, 2019 | Case-control: infants aged 1–12 months seen at gastroenterology clinic with suspected CMPA and who started CMFD for persistent GI symptoms & healthy controls | Pre-specified: score ≥12; ROC-determined: score ≥9 | Paediatrician | Response to CMFD | Change in CoMiSS™ score of symptomatic infants at presentation to a score below the median of control population after 2–4 weeks CMFD |
| Selbuz, 2020 | Infants aged 0–12 months attending paediatric outpatient clinic with a CoMiSS™ ≥12              | ROC-determined: score ≥12.5 | Paediatric gastroenterologist | OFC in infants with a change in CoMiSS™ ≥3 after CMFD | Positive if symptoms reoccur within 4 weeks of OFC                                                     |
| Sirin Kose, 2019 | Infants aged 0–10 months with a diagnosis of CMPA or HEA or both presenting to paediatric allergy clinic and with SBS ≥12 | Pre-specified: score ≥12 with reduction of 25% or score >10 with reduction of 50% | Medical supervision | OFC or skin prick test in infants <6 months | Positive if symptoms reoccur following OFC, positive skin prick with wheal >3 mm                         |
| Vandenplas, 2014 | Infants aged 0–6 months presenting with symptoms of mild-to-moderate CMPA and with SBS ≥12      | ROC-determined: score ≥12 baseline to score <6 after elimination | Physician | OFC | Positive if symptoms reoccur within 1 week of OFC                                                     |
| Zeng, 2019    | Infants aged 1–12 months seen at paediatric GI clinic for suspected CMPA with at least one symptom, including anaphylaxis, GI, respiratory and dermatological manifestations, and sIgE <5.0 kU/L | ROC-determined: score ≥5.5 | GI physician | OFC | Positive if symptoms reappear within 2 weeks of OFC                                                   |

Abbreviations: CMFD, cow’s milk-free diet; CMPA, cow’s milk protein allergy; GI, gastrointestinal; HEA, hen’s egg allergy; OFC, oral food challenge; ROC, receiver operating curve; SBS, symptom-based score; sIgE, serum IgE.

*Study used the original version of CoMiSS™ called the “symptom-based score”.*
| Study               | Score at presentation, Mean ± SD/median [IQR] | Cut-off score | Total infants in analysis | Total patients with confirmed CMPA, % prevalence | TP   | FP   | FN   | TN   | Sensitivity [95% CI]     | Specificity [95% CI] | QUADAS-2 risk of bias and applicability ratings |
|---------------------|-----------------------------------------------|---------------|---------------------------|--------------------------------------------------|------|------|------|------|--------------------------|----------------------|------------------------------------------------|
| **CMPA questionnaire** |                                               |               |                           |                                                   |      |      |      |      |                          |                      |                                                |
| Gibbons, 2012       | 10.7 ± 4.3 cases 3.7 ± 2.8 controls           | ≥6 (max 24)   | 84                        | 51% (43/84)                                      | 38   | 12   | 5    | 29   | 0.88 [0.75, 0.96]         | 0.71 [0.54, 0.84]    | ![PS IT RS FT](PS IT RS FT)                     |
| Gibbons, 2012       | 8.7 ± 3.3 cases 2.3 ± 2.2 controls            | ≥ 6 (max 15)  | 84                        | 51% (43/84)                                      | 38   | 3    | 9    | 38   | 0.81 [0.67, 0.91]         | 0.93 [0.80, 0.98]    | ![PS IT RS FT](PS IT RS FT)                     |
| **CoMiSS™**         |                                               |               |                           |                                                   |      |      |      |      |                          |                      |                                                |
| At presentation     |                                               |               |                           |                                                   |      |      |      |      |                          |                      |                                                |
| Prasad, 2018b       | 16.2 ± 6.8                                     | ≥12           | 83                        | 84% (70/83)                                      | 55   | 8    | 15   | 5    | 0.79 [0.67, 0.87]         | 0.38 [0.14, 0.68]    | ![PS IT RS FT](PS IT RS FT)                     |
| Salvatore, 2019     | 8 [2–16]                                      | ≥9            | 47                        | 40% (19/47)                                      | 16   | 4    | 3    | 24   | 0.84 [0.60, 0.97]         | 0.86 [0.67, 0.96]    | ![PS IT RS FT](PS IT RS FT)                     |
| Salvatore, 2019     | 8 [2–16]                                      | ≥12           | 47                        | 40% (19/47)                                      | 7    | 2    | 12   | 26   | 0.37 [0.16, 0.62]         | 0.93 [0.76, 0.99]    | ![PS IT RS FT](PS IT RS FT)                     |
| Selbuz, 2020c       | 13.6 ± 1.9                                     | ≥12.5         | 168                       | 54% (91/168)                                     | –    | –    | –    | –    | –                        | –                    | ![PS IT RS FT](PS IT RS FT)                     |
| Sirin Kose, 2019d   | 13 [11–16]                                    | ≥10           | 49                        | 100% (49/49)                                     | 43   | –    | 6    | –    | 0.88 [0.75, 0.95]         | Not estimable        | ![PS IT RS FT](PS IT RS FT)                     |
| Sirin Kose, 2019d   | 13 [11–16]                                    | ≥12           | 49                        | 100% (49/49)                                     | 34   | –    | 15   | –    | 0.69 [0.55, 0.82]         | Not estimable        | ![PS IT RS FT](PS IT RS FT)                     |
| Vandenplas, 2014d   | 13.65 ± 1.5                                    | ≥12           | 84                        | 69% (58/84)                                      | 58   | –    | 26   | –    | 0.69 [0.58, 0.79]         | Not estimable        | ![PS IT RS FT](PS IT RS FT)                     |
| Zeng, 2019          | 7.4 ± 2.3 CMPA+ 4.1 ± 1.6 CMPA−               | ≥5.5          | 38                        | 63% (24/38)                                      | 21   | 3    | 3    | 11   | 0.88 [0.68, 0.97]         | 0.79 [0.49, 0.95]    | ![PS IT RS FT](PS IT RS FT)                     |
| **At 1 month after ED** |                                               |               |                           |                                                   |      |      |      |      |                          |                      |                                                |
| Sirin Kose, 2019d   | –                                             | ≥25% decrease after 4-week ED                  | 49                        | 100% (49/49)                                     | 48   | –    | 1    | –    | 0.98 [0.89, 1.00]         | Not estimable        | ![PS IT RS FT](PS IT RS FT)                     |
| Sirin Kose, 2019d   | –                                             | ≥50% decrease after 4-week ED                  | 49                        | 100% (49/49)                                     | 41   | –    | 8    | –    | 0.84 [0.70, 0.93]         | Not estimable        | ![PS IT RS FT](PS IT RS FT)                     |

(Continues)
THOMPSON et al.

Reported after ROC analysis, with a best diagnostic cut-off point of 9. However, the reported accuracy estimates are likely to overestimate the real performance of the score, first because response to elimination diet was defined as a decrease in the CoMiSS™ score (incorporation bias) and the cut-off of 9 was based on ROC analysis. Selbuz evaluated CoMiSS™ against OFC in infants 0–12 months with an initial CoMiSS™ ≥12. Reported sensitivity and specificity estimates based on ROC analysis have not been included. Given that all children with CoMiSS™ <12 were excluded in the selection process without follow-up, and, that children with an improvement in CoMiSS™ <3 after elimination diet did not undergo food challenge, it is unclear what use of the test the reported estimates are referring to.

Effectiveness studies using symptom scores as an outcome

Five studies reported use of CoMiSS™ as an outcome measure (see Table 4). In all studies, after symptomatic infants were administered cow’s milk-free formula, there was a significant change in CoMiSS™ score, but no minimally important difference has been established.

3.6 CMPA questionnaire

Gibbons developed and piloted a questionnaire for non-IgE-mediated CMPA.16 A search on Gibbons found no subsequent citing articles reporting further evaluation of the tool. Study characteristics and reported sensitivity and specificity test pairs are presented in Tables 2 and 3. Test accuracy of the questionnaire at presentation relative to the results of an elimination diet in a cohort of 43 infants with non-IgE-mediated CMPA and 41 healthy controls <24 months was evaluated. ROC-determined estimates for sensitivity and specificity were 88% (95% CI: 75% to 96%) and 71% (95% CI: 54% to 84%), respectively, with a best diagnostic cut-off of ≥6. Time spent on clinical deployment of the questionnaire was reported as 3 to 6 minutes.

4 DISCUSSION

Symptom scores have been suggested to improve diagnosis of CMPA in infants by aiding preliminary clinical suspicion. We found two such scores evaluated in one and fifteen studies, respectively. These comprised evaluations of test accuracy, aspects of the development and piloting of CoMiSS™ and effectiveness studies using CoMiSS™ as an outcome measure. No end-to-end studies investigating the long-term outcomes of infants or economic evaluations were identified. Estimated sensitivity and specificity of the two symptom scores ranged from 37 to 98% and 38 to 93%, respectively.

Although we identified a number of studies evaluating the accuracy of the scores, especially CoMiSS™, we are unable to report valid estimates of their sensitivity and specificity. All studies were deemed high risk of bias, were of limited applicability and
produced highly heterogeneous results. Succeeding our decision not to pool the results given the extensive heterogeneity, one study evaluating the predictive value of CoMiSS™ after elimination diet using pooled data from three clinical trials was excluded from this review.

The accuracy of a test is dependent on multiple variables. Factors such as patient profile, previous tests and the skills and experience of the test operator could have significant impact on the performance of the test and are likely to be setting dependent. Different levels of accuracy and balance between false-positive and false-negative rate are required depending on the test's role in the diagnostic pathway. Therefore, the best approach in test evaluation is to first define the role of the test and clarify its value proposition. Given the above, the limitations of the studies included in the review could be summarised as follows.

The role of symptom scores in the diagnostic work-up of CMPA needs clarification. The CMPA questionnaire is labelled an 'important tool in determining CMPA in infants', but how it is to be used remains uncertain. CoMiSS™ is defined as an 'awareness tool' and is not intended for diagnosis of CMPA or as a replacement for OFC. Ruling out CMPA, however, is essential to avoid unnecessary referrals to allergy clinic. The tools could be used at presentation to help clinicians make a structured assessment of unexplained symptoms suggestive of CMPA. Applying the score could lead to a negative result (ruling out CMPA) which requires high sensitivity and negative predictive value; or to further tests, such as elimination diet followed by reintroduction or the need for an OFC.

Alternatively, symptom scores could be used to monitor symptoms during elimination diet or as a measure of change between baseline

| Study                  | Research objectives                                      | CoMiSS™ outcome measure | Results (N = infants included in analysis) |
|------------------------|---------------------------------------------------------|--------------------------|------------------------------------------|
| Dupont, 2016           | Measure tolerance of eHCF in infants 1–12 months with confirmed CMPA | At inclusion; day 14    | N = 30: mean reduction in CoMiSS™ score from baseline to 14 days (95% CI): -4.2 (2.8, 5.6) Mean CoMiSS™ at inclusion <12 Statistically significant reductions were reported for regurgitation and crying scores |
| Rossetti, 2019         | Evaluate hypoallergenicity of TeHCF in infants 1–36 months with suspected or diagnosed CMPA | At inclusion; day 7, day 45, day 90 | N=29: Mean reduction in CoMiSS™ score from baseline to 7 days (95% CI): 0.7 (−0.8, 2.2) Maximum score of 6 in 3/29 infants Reductions in crying scores were reported |
| Vandenplas, 2013       | Measure efficacy of eWH and eCH in infants 0.5–6 months with suspected CMPA | Before challenge; at inclusion; 1, 2, 4, 6, 8, 10 months | N = 116; mean SBS at inclusion 13.6 ± 2.2 in eWH and 13.8 ± 1.5 in eCH, SBS decreased by −8.42 in eWH and −7.8 in eCH groups at 1 month. Decrease to <2 in both groups by 10 months Mean change from eWHeWH eCH baseline (95% CI)** 8.4 (7.0, 9.9) 7.8 (6.4, 9.2) Significant reductions in all symptom scores were reported for both groups |
| Vandenplas, 2014       | Measure efficacy of eRHF in infants 0–6 months with suspected CMPA | Before challenge; at inclusion; 1, 3, 6 months | N = 38; mean reduction in SBS score from baseline to 1 month (95% CI): 10 (8.25, 11.7) Decrease to <2 by 6 months Report statistically significant reductions in regurgitation, crying, stool and dermatological scores |
| Vandenplas, 2014       | Evaluate hypoallergenicity of thickened versus non-thickened eCH in infants <6 months with suspected CMPA | At inclusion; 1 month    | N = 72; mean reduction in SBS score from baseline to 1 month (95% CI): 7.4 (6.1, 8.7). Report statistically significant reductions in regurgitation and crying scores |

Note: Note that in all studies, after symptomatic infants were administered cow's milk-free formula, there was a significant change in CoMiSS™ score, but no minimally important difference has been established.
Abbreviations: CMPA, cow's milk protein allergy; eCH, extensive casein hydrolysate; eHCF, extensively hydrolysed casein-based formula; eRHF, extensive rice hydrolysate formula; eWHeWH, extensive whey hydrolysate; SBS, symptom-based score (early version of CoMiSS™);TeHCF, thickened extensively hydrolysed casein-based formula.

**Note that the standard deviation for mean change was not reported by Rossetti, and so it was imputed assuming the standard deviation for mean change as reported by Dupont.

*Note that the standard deviations for mean change were not reported by Vandenplas 2013 and 2014a and so they were imputed assuming the standard deviation for mean change as reported by Vandenplas 2014b.
and follow-up. Some authors reported that a CoMiSS™ of <6 after 1 month of elimination diet might be more predictive of CMPA compared to baseline scores alone.11,23 A significant proportion of parents (~20%) declined OFC once symptoms became less severe or completely disappeared. We know from epidemiological studies, such as EuroPrevall, that false positives are possible with elimination diet and that elimination diet alone is not an acceptable diagnostic pathway.32 Where rechallenging is declined, the change in symptom score could be used to provide reassurance to parents that the diagnosis is correct and that a referral to allergy clinic for paediatric allergy dietitian input is in fact necessary.

In the UK, clinicians are guided by the NICE CG116 guideline when presented with infants suspected of food allergy.33 As with other national and international guidelines, including DRACMA,2 it is recommended that an allergy-focused history be taken to establish whether or not the symptoms are likely to be caused by CMPA, to decide on what type of CMPA, and whether IgE tests should be carried out prior to elimination diet or OFC. It is therefore important to consider how results from symptom scores would be used relative to the clinical information collected. For example, family history of atopy may be a significant indicator of allergy in children, and sIgE or SPT as detailed in the NICE and DRACMA guidelines.2,4,33

As outlined above, it is important to establish the type of CMPA when making a diagnosis, and thus, a good history is key. IgE-mediated CMPA reactions present immediately with recognised symptoms such as angioedema and anaphylaxis, while non-IgE-mediated CMPA is more difficult to distinguish from other childhood diseases, often presenting with generalised symptoms involving the skin (e.g. eczema), gastrointestinal system (e.g. diarrhoea, colic, vomiting and constipation) and respiratory system (e.g. wheeze) and can be delayed up to 48 hours after exposure.3,4 The CMPA questionnaire focuses on non-IgE-mediated CMPA alone, whilst CoMiSS™ appears to focus on both IgE-mediated and non-IgE-mediated CMPA, with inclusion of urticaria which is more commonly associated with IgE-mediated disease. Covering both conditions likely increases the estimated diagnostic accuracy of CoMiSS™ but may give rise to higher rates of false positives in the context of non-IgE-mediated CMPA.

Once a clear role for symptom scores is defined, and the benefit over current practice is proven, they should be rigorously evaluated with respect to patient outcomes and cost-effectiveness compared to alternative diagnostic-treatment pathways. Practical application of symptom scores for CMPA diagnosis in conjunction with other clinical methods should be assessed over extended timeframes via longitudinal studies. These should report: the recruitment of relevant patients under a clinical-testing agreement; skills and experience of clinicians running the test; provisions to deal with attrition bias (most studies reported high dropout rates); use of DBPCFC as a reference standard, which despite some conflicting evidence, is the optimum reference standard test for confirmation of CMPA in children.34

The strengths of this review include the use of internationally recommended methods for study identification and methodological quality assessment and a pre-specified protocol was registered on PROSPERO. The main limitation was the poor reporting and quality of included studies and that too few, heterogeneous studies were identified to perform meta-analysis. We did not perform independent and duplicate data extraction given limited resources.

5 | CONCLUSIONS

Just two tools for the diagnosis of CMPA were identified. The evaluations of these were either at high risk of bias or failed to address key issues such as clinical and cost-effectiveness. Estimates of accuracy offered so far should therefore be interpreted extremely cautiously. Rigorous research based on a well-defined role for the tools and free of potential conflicts of interest is urgently required.

ACKNOWLEDGEMENTS

We would like to acknowledge the contributions of Professor Stuart Logan at University of Exeter Medical School and Flamingos Allergy Toddler Group in developing this systematic review.

CONFLICTS OF INTEREST

Authors have no conflicts of interest.

AUTHOR CONTRIBUTION

Georgina Thompson: Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Writing-original draft (lead); Writing-review & editing (lead). Zhivko Zhelev: Conceptualization (equal); Data curation (supporting); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Writing-original draft (equal); Writing-review & editing (equal). Jaime Peters: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Writing-original draft (equal); Writing-review & editing (equal). Sara Khalid: Data curation (equal); Investigation (equal); Writing-original draft (supporting); Writing-review & editing (supporting). Simon Briscoe: Conceptualization (equal); Investigation (supporting); Methodology (equal); Writing-review & editing (equal). Liz Shaw: Conceptualization (equal); Methodology (equal); Writing-review & editing (equal). Michael Nunns: Conceptualization (equal); Methodology (equal); Writing-review & editing (equal). Sian Ludman: Conceptualization (equal); Writing-review & editing (equal). Christopher Hyde: Conceptualization (equal); Formal analysis (supporting); Methodology (equal); Supervision (lead); Writing-original draft (equal); Writing-review & editing (equal).

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/pai.13537.

ORCID

Georgina Thompson https://orcid.org/0000-0001-9800-7160

AUTHOR CONTRIBUTION

Georgina Thompson: Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Writing-original draft (lead); Writing-review & editing (lead). Zhivko Zhelev: Conceptualization (equal); Data curation (supporting); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Writing-original draft (equal); Writing-review & editing (equal). Jaime Peters: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Writing-original draft (equal); Writing-review & editing (equal). Sara Khalid: Data curation (equal); Investigation (equal); Writing-original draft (supporting); Writing-review & editing (supporting). Simon Briscoe: Conceptualization (equal); Investigation (supporting); Methodology (equal); Writing-review & editing (equal). Liz Shaw: Conceptualization (equal); Methodology (equal); Writing-review & editing (equal). Michael Nunns: Conceptualization (equal); Methodology (equal); Writing-review & editing (equal). Sian Ludman: Conceptualization (equal); Writing-review & editing (equal). Christopher Hyde: Conceptualization (equal); Formal analysis (supporting); Methodology (equal); Supervision (lead); Writing-original draft (equal); Writing-review & editing (equal).

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/pai.13537.

ORCID

Georgina Thompson https://orcid.org/0000-0001-9800-7160
REFERENCES

1. Grimshaw KE, Bryant T, Oliver EM, et al. Incidence and risk factors for food hypersensitivity in UK infants: results from a birth cohort study. Clin Transl Allergy. 2016;6(1):1.

2. Fiocchi A, Schunemann H, Anagnostou I, et al. The global impact of the DRACMA guidelines cow’s milk allergy clinical practice. World Allergy Organ J. 2018;11(1):2.

3. Walsh J, Meyer R, Shah N, Queckett J, Fox AT. Differentiating milk allergy (IgE and non-IgE mediated) from lactose intolerance: understanding the underlying mechanisms and presentations. Br J Gen Pract. 2016;66(649):e609-e611.

4. Walsh J, O’Flynn N. Diagnosis and assessment of food allergy in children and young people in primary care and community settings: NICE clinical guideline. Br J Gen Pract. 2011;61(588):473-475.

5. Ruffner MA, Spergel JM. Non-IgE-mediated food allergy syndromes. Ann Allergy Asthma Immunol. 2016;117(5):452-454.

6. Cerecedo I, Zamora J, Fox M, et al. The impact of double-blind placebo-controlled food challenge (DBPCFC) on the socioeconomic cost of food allergy in Europe. J Investig Allergol Clin Immunol. 2014;24(6):418-424.

7. Lozinsky AC, Meyer R, Anagnostou K, et al. Cow’s milk protein allergy from diagnosis to management: a very different journey for general practitioners and parents. Children (Basel). 2015;2(3):317-329.

8. Sicherer SH, Sampson HA. Food allergy: a review and update on epidemiology, pathogenesis, diagnosis, prevention, and management. J Allergy Clin Immunol. 2018;141(1):41-58.

9. Vandenplas Y, Dupont C, Eigenmann P, et al. A workshop report on the efficiency of the symptom-based score in infants diagnosed with cow’s milk protein and hen’s egg allergy. Allergol Immunopathol (Madri). 2019;47(3):265-271.

10. van Tulleken C. Overdiagnosis and industry influence: how cow’s milk-related symptom scoring awareness tool in young Turkish children. J Paediatr Child Health. 2020;56(11):1799-1805.

11. Vandenplas Y, Althera Study G, Steenhout P, Grathwohl D. A pilot study on the application of a symptom-based score for the diagnosis of cow’s milk protein allergy. SAGE Open Med. 2014;2:205031141523423.

12. Zeng Y, Zhang J, Dong G, et al. Assessment of Cow’s milk-related symptom scores in early identification of cow’s milk protein allergy in Chinese infants. BMC Pediatr. 2019;19(1):191.

13. Salvatore S, Bertoni E, Bogini F, et al. Testing the cow’s milk-related symptom score (CoMissTM) for the response to a cow’s milk-free diet in infants: a prospective study. Nutrients. 2019;11(10):2402.

14. Prasad R, Venkata RSA, Ghokale P, Chakravarty P, Anwar F. Cow’s Milk-related symptom Score as a predictive tool for cow’s milk allergy in Indian children aged 0–24 months. Asia Pacific Allergy. 2018;8(4):e36.

15. Dupont C, Bradatan E, Soulaines P, Nocerino R, Berni-Canani R. Tolerance and growth in children with cow’s milk allergy fed a thickened extensively hydrolyzed casein-based formula. BMC Pediatr. 2016;16(1):96.

16. Vandenplas Y, De Greef E. Extensive protein hydrolysate formula effectively reduces regurgitation in infants with positive and negative challenge tests for cow’s milk allergy. Acta Paediatr. 2014;103(6):e243-e250.

17. Vandenplas Y, De Greef E, Hauser B. Safety and tolerance of a new extensively hydrolyzed rice protein-based formula in the management of infants with cow’s milk protein allergy. Eur J Pediatr. 2014;173(9):1209-1216.

18. Vandenplas Y, Steenhout P, Planoudis Y, Grathwohl D. Treating cow’s milk protein allergy: a double-blind randomized trial comparing two extensively hydrolysed formulas with probiotics. Acta Paediatr. 2013;102(10):990-998.

19. Rossetti D, Cucchiara S, Morace A, Leter B, Oliva S. Hypoallergenicity of a thickened hydrolyzed formula in children with cow’s milk allergy. World J Clin Cases. 2019;7(16):2256-2268.

20. Schoemaker AA, Sprikkelman AB, Grimshaw KE, et al. Incidence and natural history of challenge-proven cow’s milk allergy in European children– EuroPrevall birth cohort. Allergy. 2015;70(8):963-972.

21. Food Allergy in Under 19s: Assessment and Diagnosis (Clinical Guideline CG116). National Institute for Health and Care Excellence; 2011: https://www.nice.org.uk/guidance/cg116.

22. Soares-Weiser K, Takwoingi Y, Panesar SS, et al. The diagnosis of food allergy: a systematic review and meta-analysis. Allergy. 2014;69(1):76-86.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Thompson G, Zhelev Z, Peters J, et al. Symptom scores in the diagnosis of pediatric cow’s milk protein allergy: A systematic review. Pediatr Allergy Immunol. 2021;00:1–11. https://doi.org/10.1111/pai.13537