

Symptom scores in the diagnosis of paediatric cow's milk protein allergy: a systematic review

Georgina Thompson¹, Zhivko Zhelev¹, Jaime Peters¹, Sara Khalid², Simon Briscoe², Liz Shaw², Michael Nunns², Sian Ludman³, Christopher Hyde¹

¹ Exeter Test Group, University of Exeter Medical School, University of Exeter, Exeter, UK

² University of Exeter Medical School, University of Exeter, Exeter, UK

³ Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

Running title: Symptom scores for milk allergy

Corresponding author:

Miss Georgina Thompson

Exeter Test Group

University of Exeter Medical School

Heavitree Road

Exeter

EX1 2LU

Email. gt274@exeter.ac.uk

Word count: 3479

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/PAI.13537](https://doi.org/10.1111/PAI.13537)

This article is protected by copyright. All rights reserved

MISS GEORGINA THOMPSON (Orcid ID : 0000-0001-9800-7160)

Article type : Original Article

i. Conflicts of interest

Authors have no conflicts of interest.

ii. Financial support

GT was supported by a University of Exeter paid internship for completion of this study.

iii. Abstract and keywords

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 37-98% and 38-93%. 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

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.

iv. Main text

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 commonly associated with delayed, generalised symptoms including faltering growth and gastrointestinal and dermatological manifestations, which can take up to 48 hours to develop.^{1,4} 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.^{2,9} 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 (CoMiSSTM), 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).

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.

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', 'allergy', 'infant' and 'score' (see Supplementary file 1). We searched the World Health Organisation International Clinical Trials Registry Platform. Forward citation searches on included studies were conducted using Science Citation Index and reference lists were manually inspected.

Study selection

Two authors (GT, SK) independently reviewed the titles and abstracts, and subsequently the full texts of studies identified from the literature search. Interpretative discrepancies were resolved by discussion. Study authors were contacted as needed for clarification (see Supplementary file 2).

Data extraction and quality appraisal

Standardised data extraction form was developed and piloted (GT, ZZ) (Supplementary file 3). GT extracted all data, which were then checked by a second reviewer (SK). Three authors (GT, ZZ, 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.

Data analysis, synthesis and reporting

Symptom scores were analysed separately and compared qualitatively. Results from test accuracy studies were summarised by 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).¹⁵ Data were not pooled due to the small number of studies reporting both sensitivity and specificity, and extensive clinical and methodological heterogeneity amongst included studies.

Results

Study selection

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.

Included studies

Two symptom scores were evaluated in the 16 included studies (see Table 1). One test accuracy study from the USA evaluated a CMPA questionnaire. The remaining fifteen studies focused on CoMiSSTM 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 CoMiSSTM, six test accuracy studies, and five effectiveness studies using CoMiSSTM as an outcome measure.

Description of included symptom scores

CoMiSSTM

Cow's Milk-related Symptom Score (CoMiSSTM), 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.¹⁶ 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 zero 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 Supplementary file 4. CoMiSS™ is made available online through Nestlé Health Science.¹⁷

CMPA questionnaire

Gibbons et al. 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.¹⁶ 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 Supplementary file 5.

Methodological quality of included studies

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 was 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.

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 CMPA,²² 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 Supplementary file 6.

Funding

Nine studies acknowledged conflicts of interest with regards 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}

Main findings

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 ($p < 0.001$), regurgitation ($p = 0.001$) and eczema ($p = 0.039$) 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 ($p = 0.001$) and stool ($p < 0.001$) scores.¹⁸

Vandenplas investigated the inter-rater (Health Care Professional (HCP) vs parent) variability of CoMiSS™ in 148 Spanish infants.¹⁹ Excellent agreement with an intraclass correlation coefficient (ICC) 0.981 was reported (95% CI 0.974–0.986, $p < 0.001$). 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 0.93 (95% CI 0.90–0.96; $p < 0.001$), but poorer between the HCP and parents on Day one versus Day 2, ICC 0.53 (95% CI 0.34–0.68; $p < 0.001$). 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 Table 2 and Table 3, respectively. Additional accuracy data are reported in Supplementary file 7. Forest plots are shown in Supplementary file 8.

Two studies evaluated the accuracy of CoMiSS™ to predict a positive food challenge following an elimination diet. Sirin Kose evaluated 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 $\geq 50\%$ 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 $\pm$ 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 evaluated 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 month. 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%) was 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 cutoff of 9 was based on ROC analysis. Selbuz evaluated CoMiSS™ against OFC in infants 0-12 months with an initial CoMiSS™ ≥ 12 .²² 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).²⁷⁻³¹ 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.

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.

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 37-98% and 38-93%, 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. Succeding 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 re-introduction 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 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 re-challenging 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, whilst non-IgE-mediated CMPA are 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 CoMiSSTM 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 CoMiSSTM 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, heterogenous studies were identified to perform meta-analysis. We did not perform independent and duplicate data extraction given limited resources.

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.

v. 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.

vi. References

1. Grimshaw KEC, Bryant T, Oliver EM, et al. Incidence and risk factors for food hypersensitivity in UK infants: Results from a birth cohort study. *Clinical and Translational Allergy*. 2016;6(1):1.
2. Fiocchi A, Schunemann H, Ansotegui I, et al. The global impact of the DRACMA guidelines cow's milk allergy clinical practice. *World Allergy Organization Journal*. 2018;11(1):2.
3. Walsh J, Meyer R, Shah N, Quekett 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. *Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology*. 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. *Journal of investigational allergology & clinical immunology*. 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. *Journal of Allergy and Clinical Immunology* 2018;141(1):41-58.
9. Venter C, Arshad SH. Guideline fever: an overview of DRACMA, US NIAID and UK NICE guidelines. *Curr Opin Allergy Clin Immunol*. 2012;12(3):302-315.
10. van Tulleken C. Overdiagnosis and industry influence: how cow's milk protein allergy is extending the reach of infant formula manufacturers. *BMJ*. 2018;363:k5056.
11. Vandenplas Y, Dupont C, Eigenmann P, et al. A workshop report on the development of the Cow's Milk-related Symptom Score awareness tool for young

- children. *Acta Paediatrica, International Journal of Paediatrics*. 2015;104(4):334-339.
12. Deeks J, Gatsonis Ce. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. 1.0.0 ed: The Cochrane Collaboration; 2013.
13. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.
14. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*. 2019;170(1):51-58.
15. *Review Manager (RevMan)* [computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre: The Cochrane Collaboration; 2014.
16. Gibbons TE, Patil SN, Frem JC, Smith C, Wakwe J, Swearingen CJ. Non-IgE-mediated cow milk allergy is linked to early childhood clusters of commonly seen illnesses: a pilot study. *Clinical pediatrics*. 2012;51(4):337-344.
17. The Cow's Milk-related Symptom Score CoMiSS(TM). In: Nestlé Health Science 2015: NestleHealthScience.com/CoMiSS.
18. Bigorajska K, Filipiak Z, Winiarska P, et al. Cow's Milk-Related Symptom Score in Presumed Healthy Polish Infants Aged 0-6 Months. *Paediatric Gastroenterology, Hepatology & Nutrition*. 2020;23(2):154-162.
19. Vandenplas Y, Carvajal E, Peeters S, et al. The Cow's Milk-Related Symptom Score (CoMiSS(TM)): Health Care Professional and Parent and Day-to-Day Variability. *Nutrients*. 2020;12(2):438.
20. Vandenplas Y, Salvatore S, Ribes-Koninckx C, Carvajal E, Szajewska H, Huysentruyt K. The cow milk symptom score (CoMiSSTM) in presumed healthy infants. *PLoS ONE*. 2018;13(7):e0200603.
21. Sirin Kose S, Atakul G, Asilsoy S, Uzuner N, Anal O, Karaman O. The efficiency of the symptom-based score in infants diagnosed with cow's milk protein and hen's egg allergy. *Allergologia et Immunopathologia*. 2019;47(3):265-271.
22. Selbuz SK, Altuntaş C, Kansu A, et al. Assessment of cows milk-related symptom scoring awareness tool in young Turkish children. *Journal of Paediatrics and Child Health*. 2020;56(11):1799-1805.

23. 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 medicine*. 2014;2:2050312114523423.
24. 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 Pediatrics*. 2019;19(1):191.
25. Salvatore S, Bertoni E, Bogni F, et al. Testing the cow's milk-related symptom score (CoMiss™) for the response to a cow's milk-free diet in infants: A prospective study. *Nutrients*. 2019;11(10):2402.
26. 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.
27. Dupont C, Bradatan E, Soulaïnes 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 Pediatrics*. 2016;16(1):96.
28. 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 Paediatrica, International Journal of Paediatrics*. 2014;103(6):e243-e250.
29. 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. *European Journal of Pediatrics*. 2014;173(9):1209-1216.
30. 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 Paediatrica, International Journal of Paediatrics*. 2013;102(10):990-998.
31. Rossetti D, Cucchiara S, Morace A, Leter B, Oliva S. Hypoallergenicity of a thickened hydrolyzed formula in children with cow's milk allergy. *World journal of clinical cases*. 2019;7(16):2256-2268.
32. Schoemaker AA, Sprickelman 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.

- 511 33. Food allergy in under 19s: assessment and diagnosis (Clinical guideline CG116).
512 In: National Institute for Health and Care Excellence 2011:
513 <https://www.nice.org.uk/guidance/cg116>.
514 34. Soares-Weiser K, Takwoingi Y, Panesar SS, et al. The diagnosis of food allergy:
515 a systematic review and meta-analysis. *Allergy*. 2014;69(1):76-86.
516

517 **vii. Tables**

518

519 **Table 1:** Summary of included studies.

520

Study	Country	Study type	Symptom score evaluation	Sample size (n)	Reported conflicts of interest
Bigorajska, 2020	Poland	Observational	Determine age-related CoMiSS™ scores	226	None
Dupont, 2016	United Kingdom	Observational	CoMiSS™ as an outcome measure for symptom resolution	30	Funded by industry
Gibbons, 2012	United States	Two gate DTA [†]	Test accuracy of CMPA questionnaire	84	None
Prasad, 2018	India	Single gate DTA [‡]	Test accuracy of CoMiSS™	83	Non-financial contributions from industry
Rossetti, 2019	Italy	Observational	CoMiSS™ as an outcome measure for symptom resolution	29	None
Salvatore, 2019	Italy	Single gate DTA [‡]	Test accuracy of CoMiSS™	47	None
Selbuz, 2020	Turkey	Single gate DTA [‡]	Test accuracy of CoMiSS™	168	None

Sirin Kose, 2019	Turkey	Single arm of two gate DTA [§]	Test accuracy of SBS	49	None
Vandenplas, 2013	Belgium	RCT	SBS as an outcome measure for symptom resolution	85	Unrestricted grant from Nestlé
Vandenplas, 2014	Belgium	Single gate DTA [‡]	Test accuracy of SBS	85	Financial support from Nestlé
Vandenplas, 2014	Belgium	Observational	SBS as an outcome measure for symptom resolution	38	Financial support and formula milk from United Pharmaceuticals
Vandenplas, 2014	Allar Study Group: Belgium, Greece, Kuwait, Lebanon, Slovenia	RCT	SBS as an outcome measure for symptom resolution	72	Financial support and formula milk from United Pharmaceuticals
Vandenplas, 2015	Belgium	Development	Workshop report on SBS	-	Unrestricted grant from Nestlé
Vandenplas, 2018	Belgium, Italy, Poland, Spain	Observational	Determine age-related CoMiSS™ scores	413	Employment with Nestlé and pharmaceutical companies

Vandenplas, 2019	Spain, Belgium	Observational	Inter-rater reliability and day-to-day variability of CoMiSS™ scores	220	Employment with Nestlé and pharmaceutical companies
Zeng, 2019	China	Single gate DTA [‡]	Test accuracy of CoMiSS™	38	None

[†] defined as a study in which there are two sets of inclusion criteria; one set of inclusion criteria to recruit participants suspected of the target condition, and a second set of inclusion criteria to recruit healthy participants, or those with an alternative diagnosis. Both sets of participants undergo the index test.

[‡] defined as a study in which there is a single set of inclusion criteria to recruit participants that undergo both the index test and reference standard.

[§] two gate DTA study where only the results from the participants with the target condition are reported.

CMFD – cow's milk free diet

CMPA – cow's milk protein allergy

DTA – diagnostic test accuracy

eCH – extensive casein hydrolysate

eHCF – extensively hydrolysed casein-based formula

eRHF – extensive rice hydrolysate formula

eWH – extensive whey hydrolysate

HEA – hen egg allergy

SBS – symptom-based score (early version of CoMiSS™)

TeHCF – thickened extensively hydrolysed casein-based formula

Table 2: Summary of characteristics of diagnostic test accuracy studies on CMPA questionnaire and CoMiSS™.

Study	Inclusion criteria and selection	Index test cut-off	Index test examiner	Reference standard(s)	Definition of CMPA by reference standard
<i>CMPA questionnaire</i>					
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
<i>CoMiSS™ (†or its earlier SBS version)</i>					
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 gastro-enterologist	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% <i>or</i> 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 >3mm
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

CMFD – cow's milk free diet

CMPA – cow's milk protein allergy

HEA – hen's egg allergy

GI – gastrointestinal

OFC – oral food challenge

ROC – receiver operating curve

sIgE – serum IgE

SBS – symptom-based score

526

527

528 **Table 3.** Summary of test accuracies reported by diagnostic test accuracy studies on CMPA
 529 questionnaire and CoMiSS™.
 530

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	
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	
CoMiSS™ (†or its earlier SBS version)															
At presentation															
Prasad, 2018 [‡]	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	
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	

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]	<table><tr><td>PS</td><td>IT</td><td>RS</td><td>FT</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	PS	IT	RS	FT								
PS	IT	RS	FT																				
Selbuz, 2020 ^s	13.6 ± 1.9	≥ 12.5	168	54% (91/168)	-	-	-	-	-	-	<table><tr><td>PS</td><td>IT</td><td>RS</td><td>FT</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	PS	IT	RS	FT								
PS	IT	RS	FT																				
Sirin Kose, 2019 [†]	13 [11-16]	≥ 10	49	100% (49/49)	43	-	6	-	0.88 [0.75, 0.95]	Not estimable	<table><tr><td>PS</td><td>IT</td><td>RS</td><td>FT</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	PS	IT	RS	FT								
PS	IT	RS	FT																				
Sirin Kose, 2019 [†]	13 [11-16]	≥ 12	49	100% (49/49)	34	-	15	-	0.69 [0.55, 0.82]	Not estimable	<table><tr><td>PS</td><td>IT</td><td>RS</td><td>FT</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	PS	IT	RS	FT								
PS	IT	RS	FT																				
Vandenplas, 2014 [†]	13.65 ± 1.5	≥ 12	84	69% (58/84)	58	-	26	-	0.69 [0.58, 0.79]	Not estimable	<table><tr><td>PS</td><td>IT</td><td>RS</td><td>FT</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	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]	<table><tr><td>PS</td><td>IT</td><td>RS</td><td>FT</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	PS	IT	RS	FT								
PS	IT	RS	FT																				
At 1-month after ED																							
Sirin Kose, 2019 [†]	-	≥ 25% decrease after 4-week ED	49	100% (49/49)	48	-	1	-	0.98 [0.89, 1.00]	Not estimable	<table><tr><td>PS</td><td>IT</td><td>RS</td><td>FT</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	PS	IT	RS	FT								
PS	IT	RS	FT																				

Sirin Kose, 2019 [†]	-	≥ 50% decrease after 4-week ED	49	100% (49/49)	41	-	8	-	0.84 [0.70, 0.93]	Not estimable	<table><tr><td>PS</td><td>IT</td><td>RS</td><td>FT</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	PS	IT	RS	FT								
PS	IT	RS	FT																				
Vandenplas, 2014 [†]	-	≥ 50% decrease after 4-week ED	85	69% (59/85)	45	11	14	15	0.76 [0.63, 0.86]	0.58 [0.37, 0.77]	<table><tr><td>PS</td><td>IT</td><td>RS</td><td>FT</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	PS	IT	RS	FT								
PS	IT	RS	FT																				

[‡] discrepancies in the reported data were identified. Presented here is the reported two-by-two data with test accuracy estimates calculated in Review Manager.

[§] although the authors report sensitivity and specificity estimates based on ROC-analysis, these estimates have not been included given that all children with CoMiSS™ < 12 were excluded in the selection process and never followed up to determine CMPA positivity.

[†] QUADAS-2 ratings presented as risk of bias in row one and applicability in row two. PS – patient selection; IT – index test; RS – reference standard; FT – flow and timing

CMPA – cow’s milk protein

ED – elimination diet

IQR – inter-quartile range

SBS – symptom-based score

SD – standard deviation

TP – true positive

TN – true negative

FP – false positive

FN – false negative

Table 4: Summary of characteristics of randomised controlled trials and observational studies using CoMiSS™ as an outcome measure of CMPA treatment effectiveness.

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. <table><tr><td>Mean change from</td><td>eWH</td><td>eCH</td></tr></table>	Mean change from	eWH	eCH
Mean change from	eWH	eCH				

			<table><tr><td>baseline (95%CI)**</td><td>8.4 (7.0, 9.9)</td><td>7.8 (6.4, 9.2)</td></tr></table>	baseline (95%CI)**	8.4 (7.0, 9.9)	7.8 (6.4, 9.2)
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.			

CMPA – cow's milk protein allergy

eCH – extensive casein hydrolysate

eHCF – extensively hydrolysed casein-based formula

eRHF – extensive rice hydrolysate formula

eWH – 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 Rosetti, 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.

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.

viii. **Figure legends**

Figure 1. Selection process for study inclusion.

ix. **Supplementary files**

Supplementary file 1: Search strategies.

Supplementary file 2: Authors contacted for additional data.

Supplementary file 3: Data extraction form.

Supplementary file 4: Cow Milk-related Symptom Score (CoMiSS™). Adapted from Nestle Health Science online CoMiSS™ resource.¹⁷

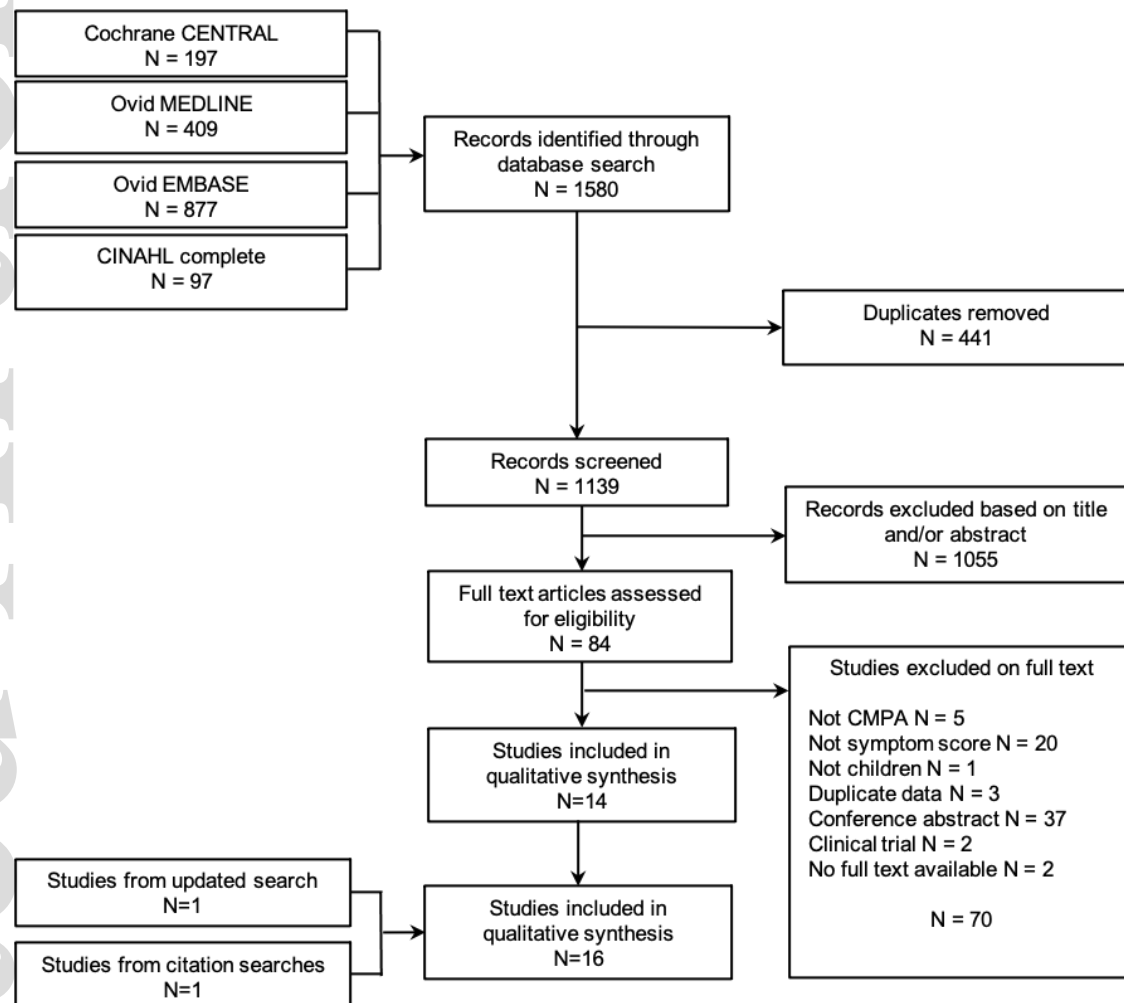
Supplementary file 5: Cow's milk protein allergy questionnaire by Gibbons et al.¹⁶

Supplementary file 6: Methodological quality of DTA studies.

Supplementary file 7. Additional test accuracy data reported by, or calculated based on data reported in, diagnostic test accuracy studies on CMPA questionnaire and CoMiSSTM. Additional data consists of area under the curve (AUC) and positive and negative likelihood ratios.

Supplementary file 8. Forest plots for test accuracy of CoMiSS tool and CMPA questionnaire.

Figure 1. Selection process for study inclusion. Databases searched from inception to 3 December 2019 (updated 14 November 2020).



1 **Supplementary file 1: Search strategies**

2 **Searches conducted 3rd December 2019**

3 Database(s): Ovid MEDLINE(R) ALL 1946 to July 26, 2019

4 Search Strategy:

# Searches	Results
1 cmpa.tw.	238
2 milk.tw.	117854
3 Milk/	54959
4 or/1-3	128570
5 (allerg* or anaphyala* or intoleran* or reaction* or symptom*).tw.	2364381
6 Hypersensitivity/	47908
7 Milk Hypersensitivity/	2395
8 "Allergy and Immunology"/	6735
9 Food Hypersensitivity/	15551
10 or/5-9	2384826
11 Checklist/	5924
12 (score* or checklist or checker or tool).tw.	1306795
13 Symptom Assessment/	4192
14 or/11-13	1311490
15 (baby or babies or child* or infant* or neonate* or newborn or schoolchild* or boy* or girl* or p?ediatric* or minor*).tw.	2261116
16 exp Infant/	1115601
17 exp child/	1865156
18 minor/	2547
19 pediatrics/	52361
20 or/15-19	3303216
21 4 and 10 and 14 and 20	409

6 Database(s): Ovid EMBASE 1974 to 2019 November 26
 7 Search Strategy:

# Searches	Results
1 cmpa.tw.	440
2 milk.tw.	129216
3 Milk/	63158
4 or/1-3	138288
5 (allerg* or anaphyala* or intoleran* or reaction* or symptom*).tw.	3090351
6 Hypersensitivity/	30312
7 Milk Hypersensitivity/	5328
8 "Allergy and Immunology"/	314014
9 Food Hypersensitivity/	26594
10 or/5-9	3373301
11 Checklist/	26594
12 (score* or checklist or checker or tool).tw.	2029070
13 Symptom Assessment/	7080
14 or/11-13	2037478
15 (baby or babies or child* or infant* or neonate* or newborn or schoolchild* or boy* or girl* or p?ediatric* or minor*).tw.	2795667
16 exp Infant/	572338
17 exp child/	1646453
18 minor/	593
19 pediatrics/	74383
20 or/15-19	3376654
21 4 and 10 and 14 and 20	877

9 Database: Cochrane CENTRAL
10 Search Strategy:

# Searches	Results
1 (cmpa):ti,ab,kw	29
2 (milk):ti,ab,kw	9977
3 Mesh descriptor: [Milk] explode all trees	2685
4 #1 or #2 or #3	10119
5 ((allerg* or anaphyala* or intoleran* or reaction* or symptom*)):ti,ab,kw	285947
6 Mesh descriptor: [Hypersensitivity] explode all trees	19517
7 Mesh descriptor: [Milk Hypersensitivity] explode all trees	234
8 Mesh descriptor: [Allergy and Immunology] explode all trees	32
9 Mesh descriptor: [Food Hypersensitivity] explode all trees	886
10 #5 or #6 or 37 or #8 or #9	296088
11 Mesh descriptor: [Checklist] explode all trees	253
12 ((score* or checklist or checker or tool)):ti,ab,kw	251020
13 Mesh descriptor: [Symptom Assessment] explode all trees	216
14 #11 or #12 or #13	251129
15 ((baby or babies or child* or infant* or neonate* or newborn or schoolchild* or boy* or girl* or p?ediatric* or minor*)):ti,ab,kw	207918
16 Mesh descriptor: [Infant] explode all trees	15714
17 Mesh descriptor: [Child] explode all trees	1203
18 Mesh descriptor: [Minor] explode all trees	8
19 Mesh descriptor: [Pediatrics] explode all trees	651
20 #15 or #16 or #17 or #18 or #19	207928
21 #4 and #10 and #14 and #20	197

12 Database: CINAHL complete
13 Search Strategy:

#	Searches	Results
1	TI (cmpa) OR AB (cmpa)	48
2	TI (milk) OR (milk)	14878
3	MH (milk)	4382
4	S1 OR S2 OR S3	16341
5	TI (allerg* or anaphyala* or intoleran* or reaction* or symptom*) OR AB (allerg* or anaphyala* or intoleran* or reaction* or symptom*)	353542
6	MH (Hypersensitivity)	8862
7	MH (Milk hypersensitivity)	607
8	MH (Food hypersensitivity)	5125
9	S5 OR S6 OR S7 OR S8	358694
10	TI (score* or checklist or checker or tool) OR AB (score* or checklist or checker or tool)	409168
11	TI (symptom assessment) OR AB (symptom assessment)	5685
12	S10 OR S11	412310
13	(baby or babies or child* or infant* or neonate* or newborn or schoolchild* or boy* or girl* or p?ediatric* or minor*) OR AB (baby or babies or child* or infant* or neonate* or newborn or schoolchild* or boy* or girl* or p?ediatric* or minor*)	611721
14	MH (infant)	154099
15	MH (child)	413489
16	MH (pediatrics)	17741
17	S13 OR S14 OR S15 OR S16	807740
18	S4 AND S9 AND S12 AND S17	97

14

15

16 **Search update conducted 14th November 2020**

17 Database(s): Ovid MEDLINE(R) ALL 1946 to July 26, 2019

18 Search Strategy:

#	Searches	Results
1	cmpa.tw.	268
2	milk.tw.	124356
3	Milk/	56712
4	or/1-3	135076
5	(allerg* or anaphyala* or intoleran* or reaction* or symptom*).tw.	2506878
6	Hypersensitivity/	48876
7	Milk Hypersensitivity/	2489
8	"Allergy and Immunology"/	6909
9	Food Hypersensitivity/	16198
10	or/5-9	2527204
11	Checklist/	6613
12	(score* or checklist or checker or tool).tw.	1434778
13	Symptom Assessment/	5327
14	or/11-13	1440449
15	(baby or babies or child* or infant* or neonate* or newborn or schoolchild* or boy* or girl* or p?ediatric* or minor*).tw.	2380148
16	exp Infant/	1148064
17	exp child/	1927660
18	minor/	2589
19	pediatrics/	54373
20	or/15-19	3441213
21	4 and 10 and 14 and 20	444

19

20

21 Database(s): Ovid EMBASE 1974 to 2019 November 26
22 Search Strategy:

#	Searches	Results
1	cmpa.tw.	481
2	milk.tw.	138095
3	Milk/	66260
4	or/1-3	147589
5	(allerg* or anaphyala* or intoleran* or reaction* or symptom*).tw.	3306055
6	Hypersensitivity/	31860
7	Milk allergy/	5587
8	Immunology/	321169
9	Food allergy/	27960
10	or/5-9	3596195
11	Checklist/	25988
12	(score* or checklist or checker or tool).tw.	2228005
13	Symptom Assessment/	8081
14	or/11-13	2237497
15	(baby or babies or child* or infant* or neonate* or newborn or schoolchild* or boy* or girl* or p?ediatric* or minor*).tw.	2969361
16	Infant/	604501
17	child/	1757894
18	minor/	660
19	pediatrics/	79123
20	or/15-19	3576833
21	4 and 10 and 14 and 20	939

23

24

25 Database: Cochrane CENTRAL
26 Search Strategy:

# Searches	Results
1 (cmpa):ti,ab,kw	31
2 (milk):ti,ab,kw	10854
3 Mesh descriptor: [Milk] explode all trees	2792
4 #1 or #2 or #3	11031
5 ((allerg* or anaphyala* or intoleran* or reaction* or symptom*)):ti,ab,kw	306756
6 Mesh descriptor: [Hypersensitivity] explode all trees	20770
7 Mesh descriptor: [Milk Hypersensitivity] explode all trees	251
8 Mesh descriptor: [Allergy and Immunology] explode all trees	34
9 Mesh descriptor: [Food Hypersensitivity] explode all trees	954
10 #5 or #6 or 37 or #8 or #9	394191
11 Mesh descriptor: [Checklist] explode all trees	266
12 ((score* or checklist or checker or tool)):ti,ab,kw	268140
13 Mesh descriptor: [Symptom Assessment] explode all trees	235
14 #11 or #12 or #13	268260
15 ((baby or babies or child* or infant* or neonate* or newborn or schoolchild* or boy* or girl* or p?ediatric* or minor*)):ti,ab,kw	219847
16 Mesh descriptor: [Infant] explode all trees	31920
17 Mesh descriptor: [Child] explode all trees	55654
18 Mesh descriptor: [Minor] explode all trees	8
19 Mesh descriptor: [Pediatrics] explode all trees	674
20 #15 or #16 or #17 or #18 or #19	219857
21 #4 and #10 and #14 and #20	325

28 Database: CINAHL complete
 29 Search Strategy:

#	Searches	Results
1	TI (cmpa) OR AB (cmpa)	63
2	TI (milk) OR (milk)	22619
3	MH (milk)	5026
4	S1 OR S2 OR S3	22643
5	TI (allerg* or anaphyala* or intoleran* or reaction* or symptom*) OR AB (allerg* or anaphyala* or intoleran* or reaction* or symptom*)	416386
6	MH (Hypersensitivity)	10030
7	MH (Milk hypersensitivity)	693
8	MH (Food hypersensitivity)	6030
9	S5 OR S6 OR S7 OR S8	422180
10	TI (score* or checklist or checker or tool) OR AB (score* or checklist or checker or tool)	482048
11	TI (symptom assessment) OR AB (symptom assessment)	6613
12	S10 OR S11	485689
13	(baby or babies or child* or infant* or neonate* or newborn or schoolchild* or boy* or girl* or p?ediatric* or minor*) OR AB (baby or babies or child* or infant* or neonate* or newborn or schoolchild* or boy* or girl* or p?ediatric* or minor*)	1038229
14	MH (infant)	173648
15	MH (child)	467411
16	MH (pediatrics)	20166
17	S13 OR S14 OR S15 OR S16	1043582
18	S4 AND S9 AND S12 AND S17	123

Supplementary file 2: Authors contacted for additional data.

Name of author	Date of contact	Nature of contact	Response
Contact through researchgate.net			
Professor Mohammad Al Enezi	19 February 2020	Request for full text. Mohammad Al Enezi et al. Safety and allergic reaction profiles of children undergoing baked milk and egg challenges: A 6-year experience from a pediatric tertiary referral center. <i>The journal of Allergy and Clinical Immunology in Practice</i> . 2017 DOI: 10.1016/j.jaip.2017.08.033.	No response
Dr R. Ibanez	19 February 2020	Request for full text Ibanez, R et al. Tolerance and efficacy of an extensively hydrolyzed formula for infants with IgE-mediated cow's milk protein allergy. JUNGLO study (part I): Extensively hydrolyzed casein formula. <i>Acta Pediátrica Española</i> 2016, 74(3):83-90.	No response
Dr S. P. Galant	19 February 2020	Request for full text. Galant, S, P et al. A potential diagnostic method for food allergy: clinical application and immunogenicity evaluation of an elemental diet. <i>American Journal of Clinical Nutrition</i> . 1977, 30(4):512-6.	No response
Dr F. Balli	19 February 2020	Request for full text. Balli, F et al. [Cow's milk protein intolerance: evaluation of immunological markers. A	No response

		preliminary study]. <i>La Pediatria medica e chirurgica: Medical and surgical pediatrics</i> . 1990, 12(3):223-8.	
Dr P. Baehler	19 February 2020	Request for full text. Baehler, P et al. Distinct patterns of cow's milk allergy in infancy defined by prolonged, two-stage double-blind, placebo-controlled food challenges. <i>Clinical & Experimental Allergy</i> . 1996, 26(3):254-61.	No response
Contact via email			
Dr Koen Huysentruyt	26 February 2020	Request for additional data. "Day to day and inter-rater variability of the Cow's Milk-related Symptom Score (CoMiSSTM)" presented at the 52nd Annual Meeting of the European Society for Paediatric Gastroenterology Hepatology and Nutrition in 2019.	Received full text Vandenplas et al. The Cow's Milk-Related Symptom Score (CoMiSSTM): Health Care Professional and Parent and Day-to-Day Variability. 2020
Dr Vincenzo Fierro	26 February 2020	Request for additional data "Hypoallergenicity and tolerance of a new amino acid-based formula in children with cow's milk allergy" presented at the 2017 World Allergy organisation symposium on hot topics in allergy: paediatric and regulatory aspects.	No response
Dr S Kaymak	4 March 2020	Request for additional data	No response

		"Assessment of Cow's milk-related symptom scoring awareness tool in Turkish infants" presented at the 51st Annual Meeting European Society for Paediatric Gastroenterology, Hepatology and Nutrition.	
Dr S K Seda	2 March 2020	Request for additional data "The effectiveness of symptom-based score on the diagnosis and follow-up of food allergy" presented at the 5th Paediatric Allergy and Asthma Meeting 2017.	No response
Professor Yvan Vandenplas	4 March 2020	Request for additional data "The cow milk symptom score (CoMiSSTM) in healthy infants" presented at 51st Annual Meeting of ESPGHAN 2018. "The cow's milk-related symptoms score (CoMiSSTM): A useful tool in the clinical setting - results from an observational study" presented at 52nd Annual Meeting of ESPGHAN 2019. "Why monitor the pH and/or impedance in the oesophagus?" presented at 10th International Congress on Pediatric Pulmonology.	No additional data available

1 **Supplementary file 3: Data extraction forms**

2

Author, year	
Country	
Funding	
Diagnostic accuracy studies	
Methods	
Study design	
Population (age/sex) & sampling	
Setting	
Inclusion criteria	
Exclusion criteria	
Index test	
Intended use	
Name of index test	
Cut-off score	
Pre-specified / ROC-determined	
Examiner number & expertise	
Time points performed	
Type of condition (IgE vs non-IgE)	
Reference standard	
Reference standard used	
Cut-off	
Time interval	
Results	
Total enrolled (sample size)	
Median (IQR) or Mean (SD) score	
Total patients with confirmed milk allergy	
Withdrawals	
Reported accuracy	
ROC analysis	
Sensitivity	
Specificity	
Two-by-two data	
True positives	
True negatives	
False positives	
False negatives	
Total included in analysis	
QUADAS-2	
Patient selection	
Risk of bias	
Applicability	
Index test	
Risk of bias	
Applicability	
Reference standard	
Risk of bias	
Applicability	
Flow and timing	
Risk of bias	
Effectiveness studies	
Research question	
Population	
Intervention(s)	
Comparator	
Outcomes	
Results	

Methodological quality (Cochrane risk of bias tool)	
Domain 1: randomisation / confounding	
Domain 2: assignment / selection	
Domain 3 (ROBINS 6): missing data	
Domain 4 (ROBINS 7): outcome measure	
Domain 5: reported result	
Domain 6 (ROBINS 3): intervention classification	
Domain 7 (ROBINS 4): dropouts	
Overall risk of bias	
Development or pilot studies	
Intended use	
Background info on symptom score	
Target population	
Methods	
Outcome	
Methodological quality (PROBAST)	
Domain 1: participants	
Risk of bias	
Applicability	
Domain 2: predictors	
Risk of bias	
Applicability	
Domain 3: outcome	
Risk of bias	
Applicability	
Analysis	
Risk of bias	
Overall risk of bias	
Overall applicability	

1 **Supplementary file 4:** Cow Milk-related Symptom Score (CoMiSS™). Adapted from
2 Nestle Health Science online CoMiSS™ resource.¹⁷

SYMPTOM	SCORE	
CRYING †	0	< 1 h/d
	1	1-1.5 h/d
	2	1.5-2 h/d
	3	2-3 h/d
	4	3-4 h/d
	5	4-5 h/d
	6	> 5 h/d
REGURGITATION	0	0-2 episodes/d
	1	3-5 of small volume
	2	> 5 episodes of > 1 coffee spoon
	3	> 5 episodes of half of the feed in < half of the feeds
	4	Continuous regurgitations of small volumes > 30 minutes after each feeding
	5	Regurgitation of half to complete volume of feed in at least half of the feeds
STOOLS (BRISTOL SCALE)	6	Regurgitation of the complete feed after each feeding
	4	Type 1 and 2 (hard stools)
	0	Type 3 and 4 (normal stools)
	2	Type 5 (soft stool)
	4	Type 6 (loose stool, if unrelated to infection)
	6	Type 7 (watery stools)
SKIN SYMPTOMS	0-6	Atopic eczema: Head-neck-trunk (absent, 0; mild, 1; moderate, 2; severe, 3) Arms-hands-legs-feet (absent, 0; mild, 1; moderate, 2; severe, 3)
	0	No urticaria
	6	Urticaria present
RESPIRATORY SYMPTOMS	0	No respiratory symptoms
	1	Slight symptoms
	2	Mild symptoms
	3	Severe symptoms

4
5 † Crying was only considered if the child was crying for one week or more, assessed by the
6 parents, without any other obvious cause.

1 **Supplementary file 5: Cow's milk protein allergy questionnaire by Gibbons et al.¹⁶**
2

QUESTION		SCORE	
Q01	How often did your child vomit in the first month?	No vomiting Less than 5 time per day More than 5 times per day	0 1 2
Q02	Were different formulas tried?	0 = No	1 = Yes
Q03	Did child gag or choke regularly on feeds?	0 = No	1 = Yes
Q04	Was your child's doctor concerned about poor weight gain in first year?	0 = No	1 = Yes
Q05	Did you consider your child to have excessive irritability or colic?	0 = No	1 = Yes
Q06	Did your child pass large painful stools? †	0 = No	1 = Yes
Q07	Did your child have blood in the stools? †	0 = No	1 = Yes
Q08	Was there always lots of mucus or "cold" in the stool? †	0 = No	1 = Yes
Q09	Did you consider your child to have diarrhoea?	0 = No	1 = Yes
Q10	Did your child have cradle cap or "scabs" in his/her hair? †	0 = No	1 = Yes
Q11	Was this treated with anything? †	0 = No	1 = Yes
Q12	Did your child have recurrent eczema or rashes? †	0 = No	1 = Yes
Q13	Was this treated with anything? †	0 = No	1 = Yes
Q14	Did your child have wheezing or persistent cough?	0 = No	1 = Yes
Q15	Was this treated with any medicines? †	0 = No	1 = Yes
Q16	Did your child have persistent congestion in the nose and eyes?	0 = No	1 = Yes
Q17	Was this treated with medicines? †	0 = No	1 = Yes
Q18	Did your child have more than 3 episodes of ear infection in the first year?	0 = No	1 = Yes
Q19	Did your child have ear tubes placed?	0 = No	1 = Yes
Q20	Was your child diagnosed with a swallow disfunction?	0 = No	1 = Yes
Q21	Are there any siblings? †	0 = No	1 = Yes
Q22	Did siblings need a change in formula as a young infant? †	0 = No	1 = Yes
Q23	Did any siblings have similar problems as a young infant?	0 = No	1 = Yes
Q24	Did mom or dad have problems as infants that required a formula change or needed goat's milk?	0 = No	1 = Yes
Q25	Were there any other unusual problems you noted in your child?	0 = No	1 = Yes
TOTAL SCORE (0-24)			

3
4 † indicates questions that could be left out of final questionnaire because of no significant
5 difference between cow's milk protein allergy and control groups.

1 **Supplementary file 6: Methodological quality of DTA studies**

2

3

4 **QUADAS-2 forms**

5

6 **Phase 1: State the review question:**

7

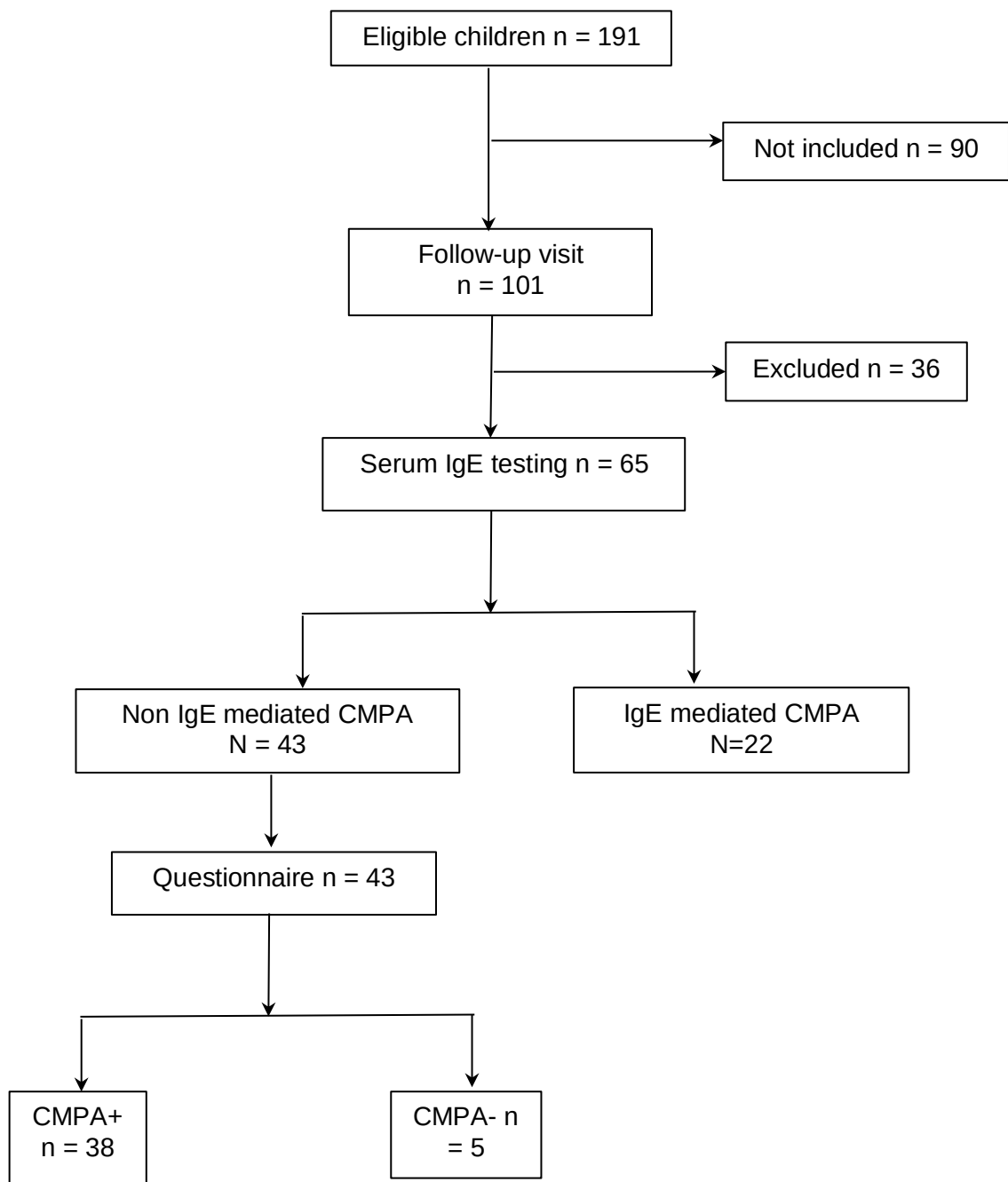
<i>Patients:</i> infants suspected of having CMPA. Studies of adults or infants already with a diagnosis of CMPA will be excluded.
<i>Index test:</i> symptom-based score / checklist providing likely probability of having CMPA.
<i>Reference standard and target condition:</i> preferred reference standard will be double-blind placebo-controlled food challenge, but studies with other reference standards will be included.

8

9

Study: Gibbons 2012

Phase 2: Flow diagram for “cases”.



20 **Phase 3: Risk of bias and applicability judgments.**
 21

DOMAIN 1: PATIENT SELECTION	
A. Risk of Bias	
Describe methods of patient selection:	
Was a consecutive or random sample of patients enrolled? No mention of consecutive/random sampling	Yes/no/ unclear
Was a case-control design avoided?	Yes/ no /unclear
Did the study avoid inappropriate exclusions? Reasons for exclusions not recorded	Yes/ no /unclear
Could the selection of patients have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting): Infants enrolled at GI clinic with suspected CMPA Presumed healthy infants enrolled at routine vaccinations	
Is there concern that the included patients do not match the review question?	CONCERN: LOW/HIGH/UNCLEAR

22

23

DOMAIN 2: INDEX TEST(S)	
If more than one index test was used, please complete for each test.	
A. Risk of Bias	
Describe the index test and how it was conducted and interpreted: Questionnaire developed by study group based on preliminary retrospective review of patients charts who were diagnosed with CMPA and a literature review defining manifestations of CMPA Yes (1 point) or No (0 points) scoring system with a total of 0-24 Questionnaire completed by healthcare professional	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes/ no /unclear
If a threshold was used, was it pre-specified?	Yes/ no /unclear
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differs from the review question?	CONCERN: LOW/HIGH/UNCLEAR

24

25

DOMAIN 3: REFERENCE STANDARD	
A. Risk of Bias	
Describe the reference standard and how it was conducted and interpreted: Non-IgE mediated disease confirmed with allergen-specific serum assay and 2-week elimination diet leading to clinically significant improvement.	
Is the reference standard likely to correctly classify the target condition? Only elimination diet used, no food challenge.	Yes/ no /unclear
Were the reference standard results interpreted without knowledge of the results of the index test?	Yes /no/unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	

Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW/HIGH/UNCLEAR
---	------------------------------

26

27

DOMAIN 4: FLOW AND TIMING	
A. Risk of Bias	
Describe any patients who did not receive the index test(s) and/or reference standard who were excluded from the 2x2 table: N=90 eligible children excluded (no reasons) N=36 excluded before serum IgE testing (no reasons) No information with regards to infants lost to follow up or those who did not want serum testing Describe the time interval and any interventions between index test(s) and reference standard: Questionnaire administered in a cross-sectional manner at the same time in both the case and cohort samples. No information with regards to when questionnaire was administered in relation to timing of the IgE serum test / inclusion of healthy infants	
Was there an appropriate interval between index test(s) and reference standard? An appropriate interval is defined as up to six weeks.	Yes/no/unclear
Did patients receive the same reference standard? Different age groups receiving different reference standard.	Yes/no/unclear
Were all patients included in the analysis?	Yes/no/unclear
Could the patient flow have introduced bias?	RISK: LOW/HIGH/UNCLEAR

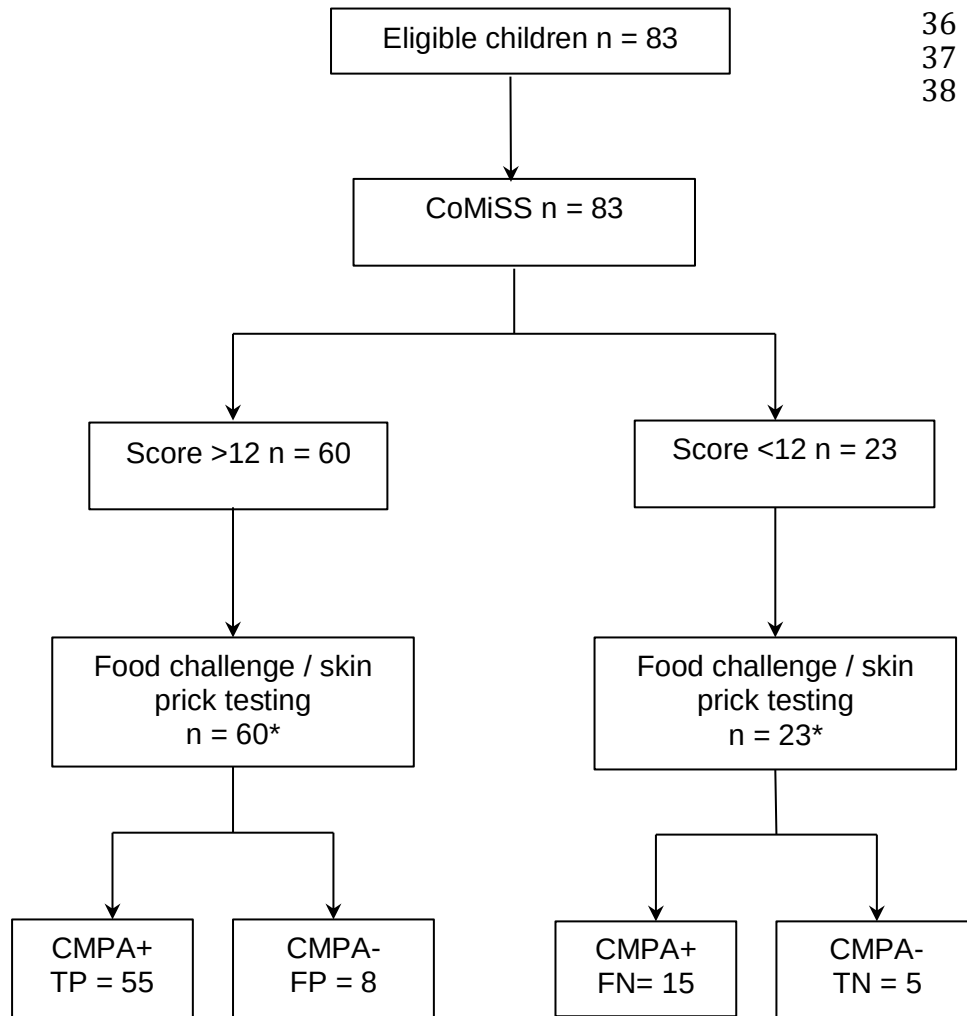
28

29

Study: Prasad 2019

Phase 2: Flow diagram.

*discrepancy in reported number of infants



39 **Phase 3: Risk of bias and applicability judgments.**
40

DOMAIN 1: PATIENT SELECTION	
A. Risk of Bias	
Describe methods of patient selection: All infants with symptoms presenting to paediatric clinic Only infants with acute diarrhoea, congenital abnormalities and established causes of other paediatric illness were excluded	
Was a consecutive or random sample of patients enrolled? No mention of consecutive/random sampling.	Yes/no/ unclear
Was a case-control design avoided?	Yes /no/ unclear
Did the study avoid inappropriate exclusions? Exclusions detailed above	Yes/no/ unclear
Could the selection of patients have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting): No prior testing in infants but they each had at least one symptom of CMPA to be included Setting in a paediatric clinic in the hospital centres included in the study	
Is there concern that the included patients do not match the review question?	CONCERN: LOW/HIGH/UNCLEAR

41

42

DOMAIN 2: INDEX TEST(S)	
If more than one index test was used, please complete for each test.	
A. Risk of Bias	
Describe the index test and how it was conducted and interpreted: CoMiSS symptom score used on all included children Score >12 positive for CMPA	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes /no/ unclear
If a threshold was used, was it pre-specified?	Yes /no/ unclear
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differs from the review question?	CONCERN: LOW/HIGH/UNCLEAR

43

44

DOMAIN 3: REFERENCE STANDARD	
A. Risk of Bias	
Describe the reference standard and how it was conducted and interpreted: Oral food challenge, skin prick test or immunoCAP test for IgE antibodies OFC carried out 15 days after elimination diet and absence of symptoms with follow-up period of up to 4 weeks. Asymptomatic children considered negative for CMPA and children whose symptoms reappeared were considered positive for CMPA Not clear who carried out the reference tests and whether the results were interpreted without knowledge of the results of the index test; potential differential verification bias (different reference standard used to verify results)	
Is the reference standard likely to correctly classify the target condition?	Yes/ no / unclear

Were the reference standard results interpreted without knowledge of the results of the index test?	Yes/no/ unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW/HIGH/UNCLEAR

45

46

DOMAIN 4: FLOW AND TIMING	
A. Risk of Bias	
Describe any patients who did not receive the index test(s) and/or reference standard who were excluded from the 2x2 table: All 83 infants included in the study had a CoMiSS score and oral food challenge test / IgE testing No exclusions with reasons are recorded for the selection process of patients No withdrawals from the study Describe the time interval and any interventions between index test(s) and reference standard: Not clear when the confirmatory diagnosis of CMPA took place after CoMiSS scoring	
Was there an appropriate interval between index test(s) and reference standard?	Yes/no/ unclear
An appropriate interval is defined as up to six weeks.	
Did patients receive the same reference standard?	Yes/ no /unclear
Different age groups receiving different reference standard.	
Were all patients included in the analysis?	Yes/no/ unclear
Discrepancies in numbers reported	
Could the patient flow have introduced bias?	RISK: LOW/HIGH/UNCLEAR

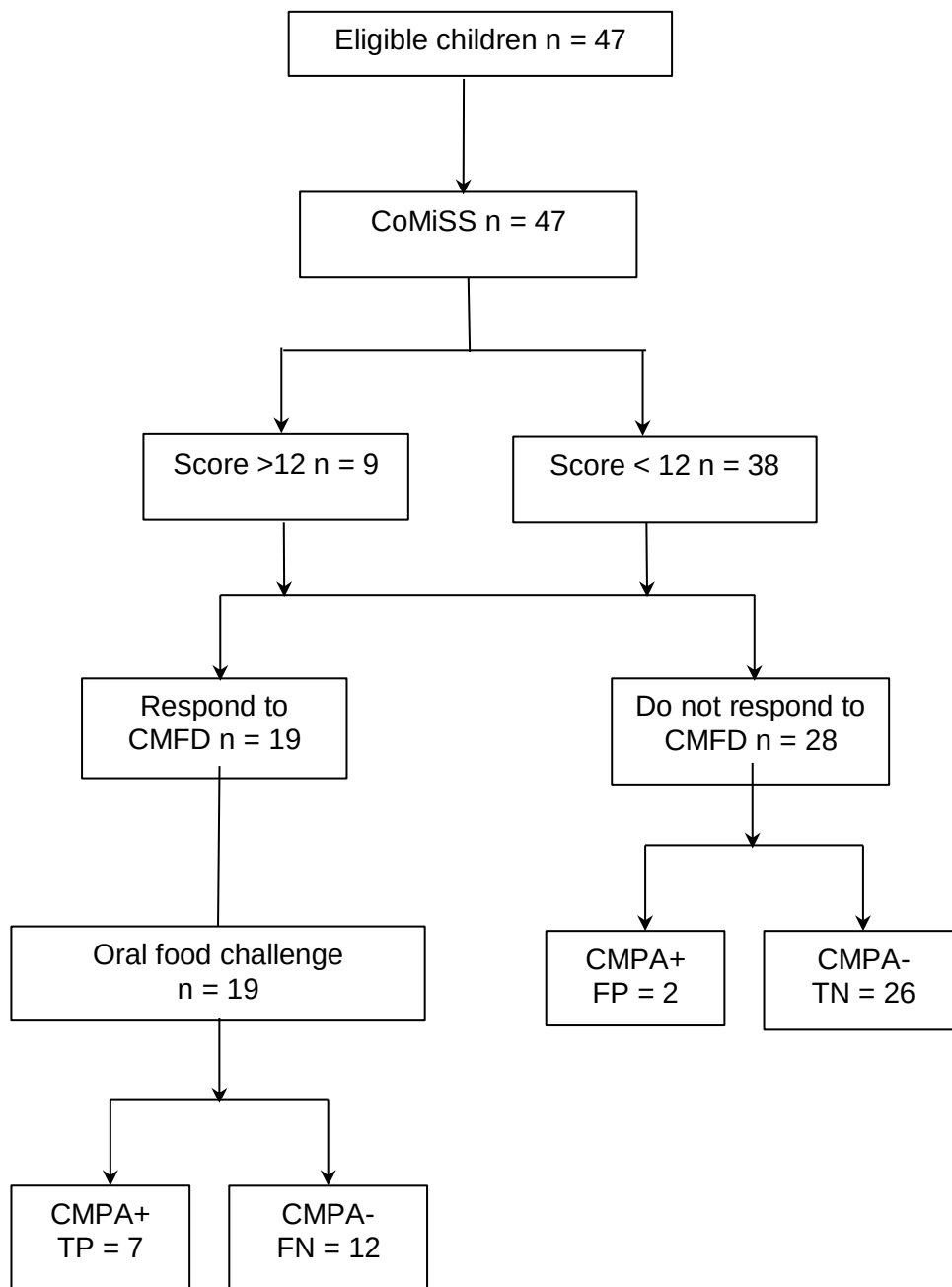
47

48

Study: Salvatore 2019

Phase 2: Flow diagram.

94 healthy controls were also included in an attempt to do a two-gate study, but these controls were not included in the overall results



60 **Phase 3: Risk of bias and applicability judgments.**
61

DOMAIN 1: PATIENT SELECTION	
A. Risk of Bias	
Describe methods of patient selection: Infants referred to paediatric gastroenterology clinic who started cow milk free diet due to persistent GI symptoms Clear exclusion criteria but reasons for exclusions not given in results Healthy controls recruited from referrals for minor trauma at the hospital, but these were used for a separate analysis	
Was a consecutive or random sample of patients enrolled? All eligible infants from one clinic were selected but no mention of consecutive or random sampling	Yes/no/ unclear
Was a case-control design avoided?	Yes /no/ unclear
Did the study avoid inappropriate exclusions? Reference standard only applied to children who responded to the cow milk free diet. 94 healthy controls were not included.	Yes/no/ unclear
Could the selection of patients have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting): Infants were not tested for CMPA prior to the study Healthcare setting appropriate and appropriate symptoms for inclusion used	
Is there concern that the included patients do not match the review question?	CONCERN: LOW/HIGH/UNCLEAR

62
63

DOMAIN 2: INDEX TEST(S)	
If more than one index test was used, please complete for each test.	
A. Risk of Bias	
Describe the index test and how it was conducted and interpreted: CoMiSS symptom score evaluated Positive score >12 and negative score <12 Conducted by paediatric gastroenterologist who was blind to CoMiSS	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes /no/ unclear
If a threshold was used, was it pre-specified?	Yes /no/ unclear
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differs from the review question?	CONCERN: LOW/HIGH/UNCLEAR

64
65

DOMAIN 3: REFERENCE STANDARD	
A. Risk of Bias	
Describe the reference standard and how it was conducted and interpreted: Allergy tests and open food challenge on all infants who commenced cow milk free diet, however, reference standard was in fact change in CoMiSS™ score of symptomatic infants at presentation to a score below the median of control population after 2-4 weeks CMFD. Performed in hospital by experienced clinicians who were blind to results of CoMiSS Positive food challenge considered with at least one symptom occurring	

Is the reference standard likely to correctly classify the target condition? Not DBPCFC	Yes/no/unclear
Were the reference standard results interpreted without knowledge of the results of the index test?	Yes/no/unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW/HIGH/UNCLEAR

66

67

DOMAIN 4: FLOW AND TIMING	
A. Risk of Bias	
Describe any patients who did not receive the index test(s) and/or reference standard who were excluded from the 2x2 table (refer to flow diagram): All n=47 children had CoMiSS score Only CMFD responding infants had oral food challenge n=19 N= 28 excluded from oral food challenge	
Describe the time interval and any interventions between index test(s) and reference standard: Within three months following inclusion	
Was there an appropriate interval between index test(s) and reference standard? Oral food challenge within 3 months of CoMiSS score	Yes/no/unclear
Did patients receive the same reference standard? All had oral food challenge and skin prick test	Yes/no/unclear
Were all patients included in the analysis?	Yes/no/unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW/HIGH/UNCLEAR

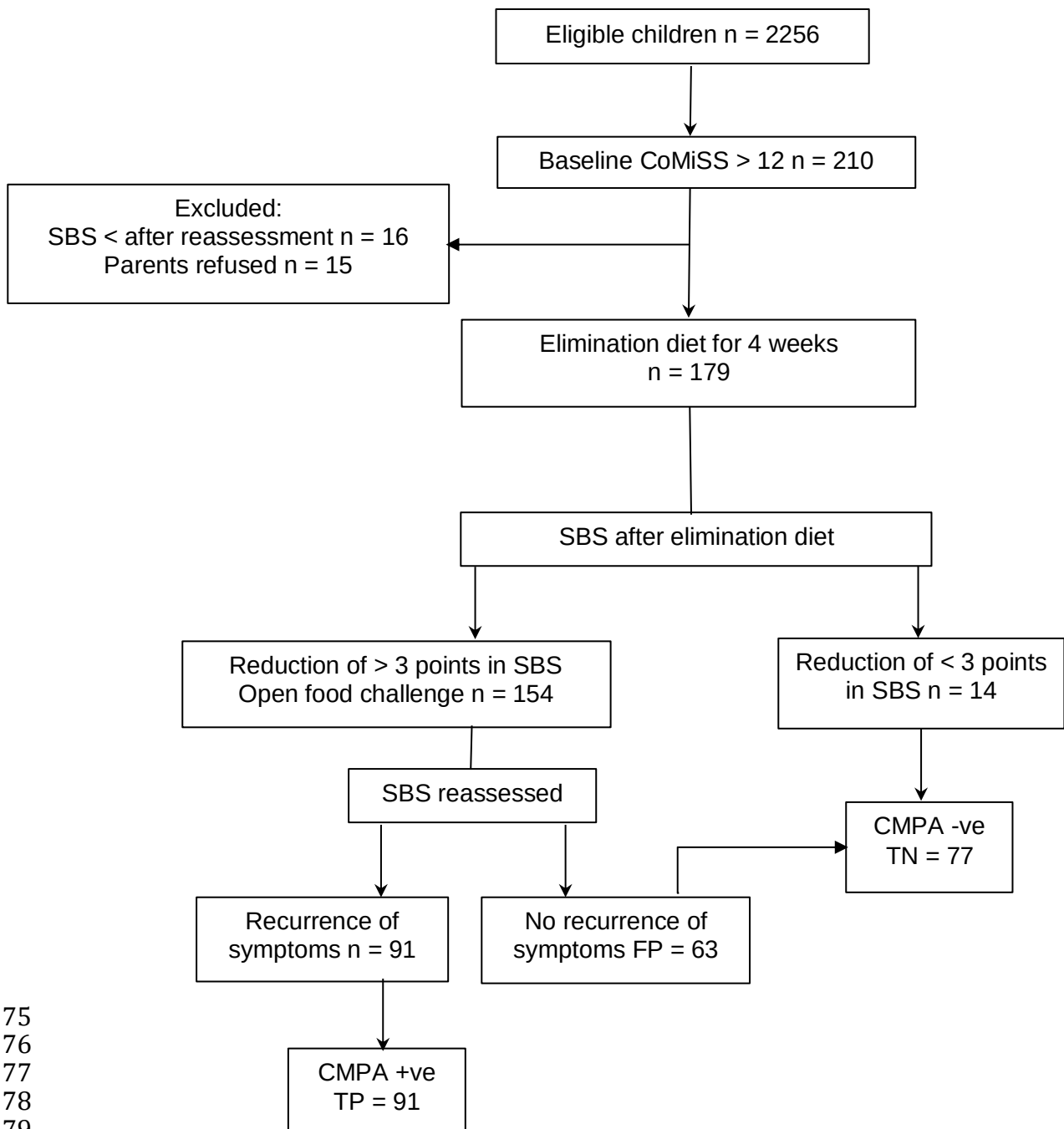
68

69

70

Study: Selbuz

Phase 2: Flow diagram for the primary study.



82 **Phase 3: Risk of bias and applicability judgments.**
83

DOMAIN 1: PATIENT SELECTION	
A. Risk of Bias	
Describe methods of patient selection: Every infant 0-1 year presenting to Paediatric outpatient department was assessed with CoMiSS. Those scoring CoMiSS > 12 on reassessment were included in the study. Infants receiving cow-milk free formula, those with known lactose intolerance, severe chronic diarrhoea, failure to thrive, multiple food allergy, blood in stool or who recently had a surgical intervention or medical treatment were excluded.	
Was a consecutive or random sample of patients enrolled? Every infant attending clinic was selected	Yes/no/unclear
Was a case-control design avoided?	Yes/no/unclear
Did the study avoid inappropriate exclusions? Children with proctocolitis excluded	Yes/no/unclear
Could the selection of patients have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting): Infants had not undergone prior testing but had a CoMiSS > 12 at inclusion. Appropriate setting and use of CoMiSS in a paediatric gastroenterology clinic	
Is there concern that the included patients do not match the review question? Patients were included based on the evaluated test. They did not have a prior suspicion of CMPA.	CONCERN: LOW/HIGH/UNCLEAR

84
85

DOMAIN 2: INDEX TEST(S)	
If more than one index test was used, please complete for each test.	
A. Risk of Bias	
Describe the index test and how it was conducted and interpreted: CoMiSS symptom score was used CoMiSS scored on three visits: at baseline, after 1-month elimination diet, and after OFC	
Were the index test results interpreted without knowledge of the results of the reference standard? The index test was not independent of the reference standard	Yes/no/unclear
If a threshold was used, was it pre-specified? Threshold of 12.5 decided via ROC-analysis	Yes/no/unclear
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differs from the review question? All infants have a baseline CoMiSS of > 12	CONCERN: LOW/HIGH/UNCLEAR

86
87

DOMAIN 3: REFERENCE STANDARD	
A. Risk of Bias	
Describe the reference standard and how it was conducted and interpreted: 4 weeks of elimination diet with ehF	

Infants with reduction in CoMiSS of < 3 points switched to aaF for 4 weeks. If unresponsive to both formulas, the infant was deemed CMPA negative. Infants with reduction in CoMiSS of > 3 points underwent open food challenge Challenges performed under medical supervision in paediatric gastroenterology clinic	
Is the reference standard likely to correctly classify the target condition? Not DBPCFC	Yes/no/unclear
Were the reference standard results interpreted without knowledge of the results of the index test? The reference standard incorporated the index test; CMPA was ruled out in 14 infants based on their CoMiSS score following elimination diet	Yes/no/unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW/HIGH/UNCLEAR

88

89

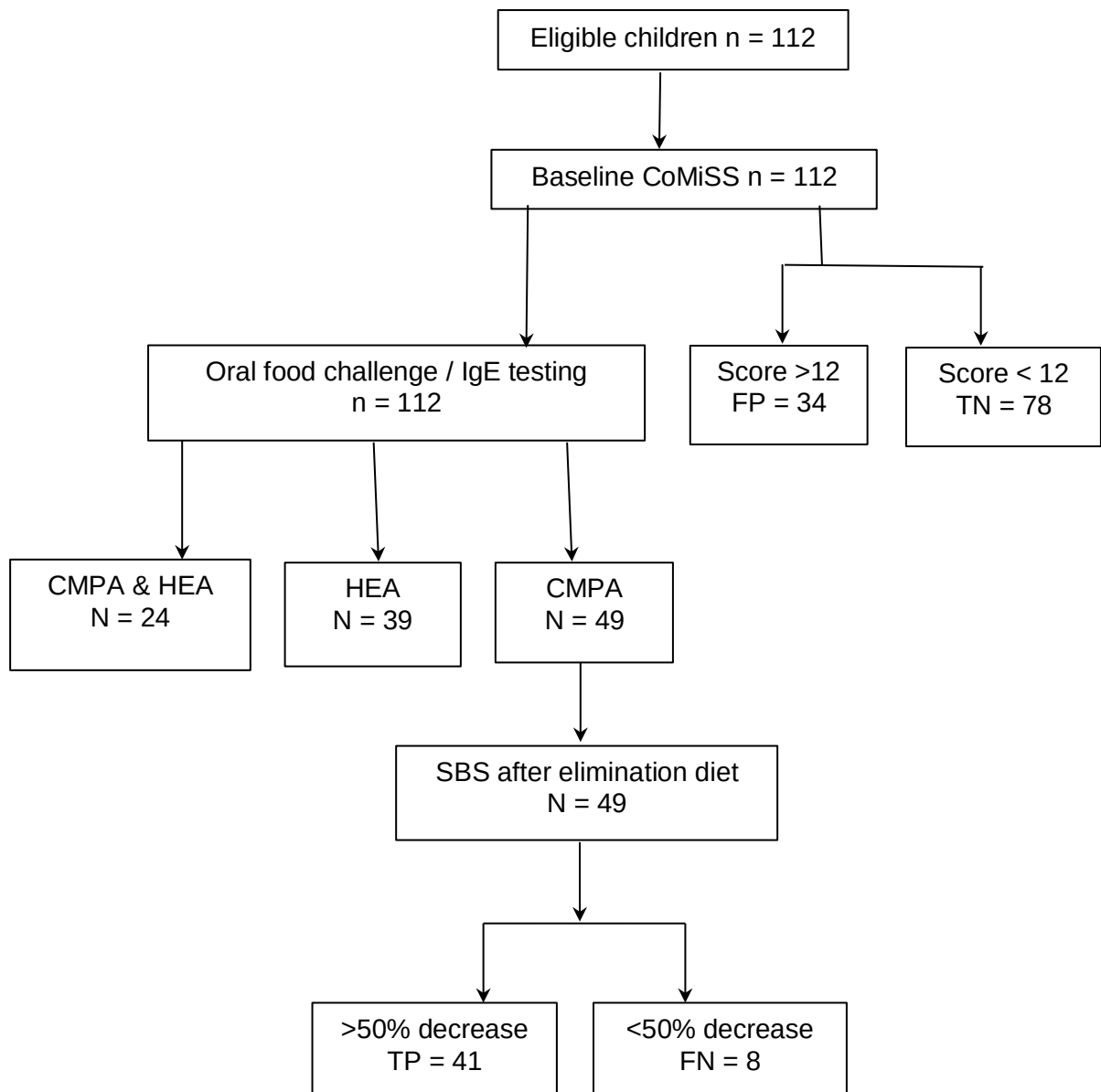
DOMAIN 4: FLOW AND TIMING	
A. Risk of Bias	
Describe any patients who did not receive the index test(s) and/or reference standard who were excluded from the 2x2 table (refer to flow diagram): Exclusions: infants CoMiSS < 12 at reassessment (n=16); parents' refusal (n=15); mothers' non-adherent to elimination diet (n=11) N=179 CoMiSS > 12 on reassessment and started elimination diet N=168 completed elimination diet and CoMiSS performed N=154 infants with reduction in CoMiSS > 3 points had oral food challenge testing; n=91 CMPA+ and n=63 CMPA- Describe the time interval and any interventions between index test(s) and reference standard: No specific interval recorded but oral food challenge conducted after 4 weeks elimination diet	
Was there an appropriate interval between index test(s) and reference standard? Roughly 4 weeks	Yes/no/unclear
Did patients receive the same reference standard? In 14 children CMPA was ruled out based on CoMiSS score after elimination diet, and did not have OFC	Yes/no/unclear
Were all patients included in the analysis?	Yes/no/unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW/HIGH/UNCLEAR

90

91

Study: Sirin Kose

Phase 2: Flow diagram.



101 **Phase 3: Risk of bias and applicability judgments.**
102

DOMAIN 1: PATIENT SELECTION	
A. Risk of Bias	
Describe methods of patient selection: Patients selected from Paediatric immunology and allergy outpatient clinic who had been diagnosed with CMPA, HEA or both Infants exhibiting symptoms of Gi or respiratory infection were excluded	
Was a consecutive or random sample of patients enrolled? Only infants with established diagnosis were included – not consecutive or random	Yes/no/unclear
Was a case-control design avoided?	Yes/no/unclear
Did the study avoid inappropriate exclusions? Reference standard was applied to all included infants Only CMPA infants were included in the accuracy data for using CoMiSS to detect infants with CMPA	Yes/no/unclear
Could the selection of patients have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting): Infants were diagnosed with CMPA prior to the study Appropriate setting and use of CoMiSS in a paediatric immunology and allergy clinic	
Is there concern that the included patients do not match the review question?	CONCERN: LOW/HIGH/UNCLEAR

103

104

DOMAIN 2: INDEX TEST(S)	
If more than one index test was used, please complete for each test.	
A. Risk of Bias	
Describe the index test and how it was conducted and interpreted: CoMiSS symptom score used at baseline and after 4-week elimination diet – the change in score post-elimination diet is used for test accuracy estimates	
Were the index test results interpreted without knowledge of the results of the reference standard? Does not specifically say whether the paediatricians were blind to CoMiSS	Yes/no/unclear
If a threshold was used, was it pre-specified?	Yes/no/unclear
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differs from the review question? If the test is to be used before elimination, then the results are not really applicable	CONCERN: LOW/HIGH/UNCLEAR

105

106

DOMAIN 3: REFERENCE STANDARD	
A. Risk of Bias	
Describe the reference standard and how it was conducted and interpreted: Oral food challenge consisting of 4 weeks elimination diet followed by skin prick testing Infants <6 months were deemed CMPA positive if their symptoms improved with 4-week maternal elimination diet; Infants <6 months deemed CMPA positive if symptoms improve with extensively hydrolyzed formula	

Challenges performed under medical supervision	
Is the reference standard likely to correctly classify the target condition? No DBPCFC	Yes/no/unclear
Were the reference standard results interpreted without knowledge of the results of the index test?	Yes/no/unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW/HIGH/UNCLEAR

107

108

DOMAIN 4: FLOW AND TIMING	
A. Risk of Bias	
Describe any patients who did not receive the index test(s) and/or reference standard who were excluded from the 2x2 table (refer to flow diagram): No exclusions in study. All of the infants are included in the analysis N=49 infants with CMPA had both CoMiSS and oral food challenge testing Describe the time interval and any interventions between index test(s) and reference standard: No specific interval recorded but oral food challenge conducted after 4 weeks elimination diet	
Was there an appropriate interval between index test(s) and reference standard? Roughly 4 weeks	Yes/no/unclear
Did patients receive the same reference standard? Breastfed infants <6 months did not undergo food challenge They had skin prick test instead	Yes/no/unclear
Were all patients included in the analysis?	Yes/no/unclear
Could the reference standard, its conduct, or its interpretation have introduced bias? Risk of differential verification bias	RISK: LOW/HIGH/UNCLEAR

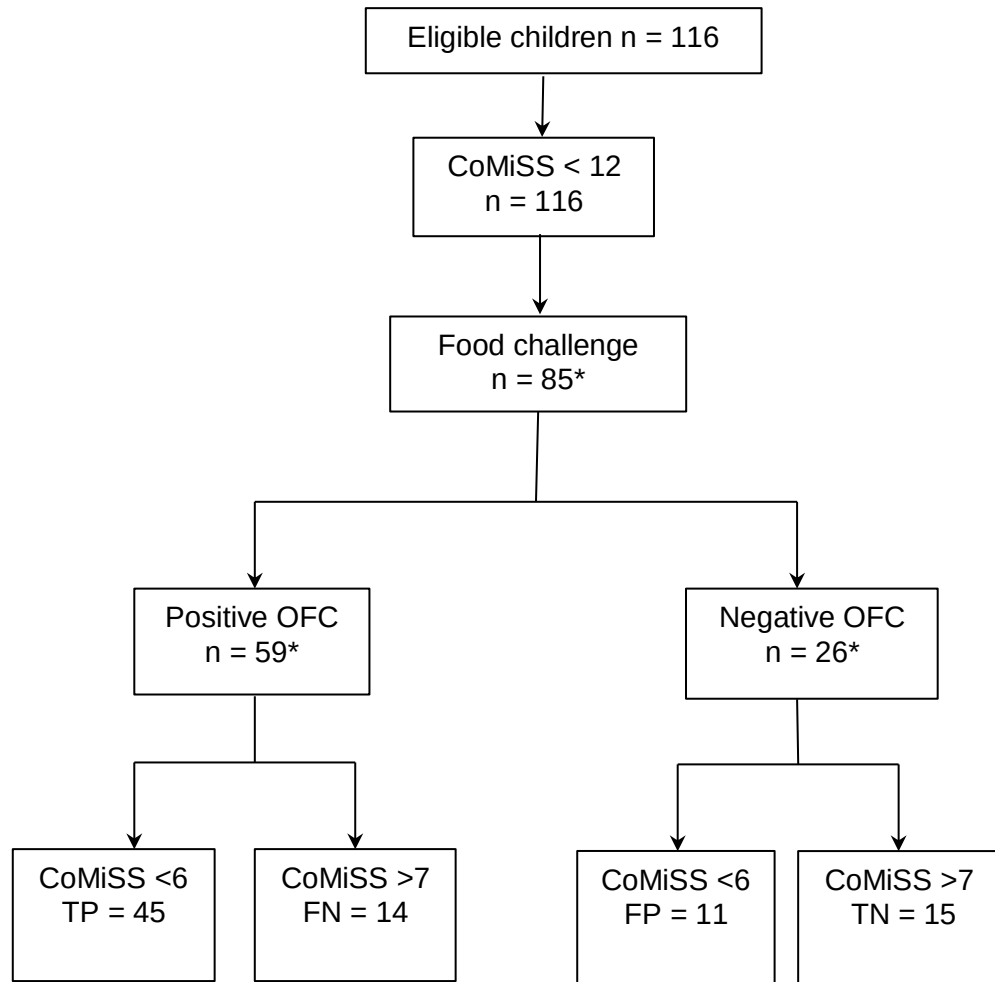
109

110

Study: Vandenplas 2014

Phase 2: Flow diagram.

*Discrepancy in reported numbers in the text and in Table 4



122 **Phase 3: Risk of bias and applicability judgments.**
123

DOMAIN 1: PATIENT SELECTION	
A. Risk of Bias	
Describe methods of patient selection: Formula fed infants presenting with symptoms of mild to moderate CMPA CoMiSS score >12 required for inclusion into the study	
Was a consecutive or random sample of patients enrolled? Only patients with CoMiSS >12 included	Yes/no/ unclear
Was a case-control design avoided?	Yes /no/ unclear
Did the study avoid inappropriate exclusions? N=31 infants did not undergo food challenge as parents refused; all infants had an SBS >12 at presentation and one child who scored SBS=5 was considered a "protocol deviation"	Yes/ no / unclear
Could the selection of patients have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting): No prior testing, setting for CoMiSS is not clear	
Is there concern that the included patients do not match the review question?	CONCERN: LOW/HIGH/UNCLEAR

124

125

DOMAIN 2: INDEX TEST(S)	
If more than one index test was used, please complete for each test.	
A. Risk of Bias	
Describe the index test and how it was conducted and interpreted: CoMiSS symptom score used at baseline and after 4-week elimination diet – the change in score post-elimination diet is used for test accuracy estimates Unclear as to who completed the CoMiSS symptom score and whether the second CoMiSS was interpreted without knowledge of the reference standard Initial score of >12 used in inclusion Hypothesized that a decrease in score of >3 points would be clinically meaningful	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes/no/ unclear
If a threshold was used, was it pre-specified?	Yes/ no / unclear
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differs from the review question? If the test is to be used before elimination, then the results are not really applicable	CONCERN: LOW/HIGH/UNCLEAR

126

127

DOMAIN 3: REFERENCE STANDARD	
A. Risk of Bias	
Describe the reference standard and how it was conducted and interpreted: Gold standard test oral food challenge to confirm diagnosis of CMPA Test conducted in hospital by qualified professionals after 4-week elimination diets Unclear whether the examiners were blinded	

Is the reference standard likely to correctly classify the target condition? Not DBPCFC	Yes/no/unclear
Were the reference standard results interpreted without knowledge of the results of the index test?	Yes/no/unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW/HIGH/UNCLEAR

128

129

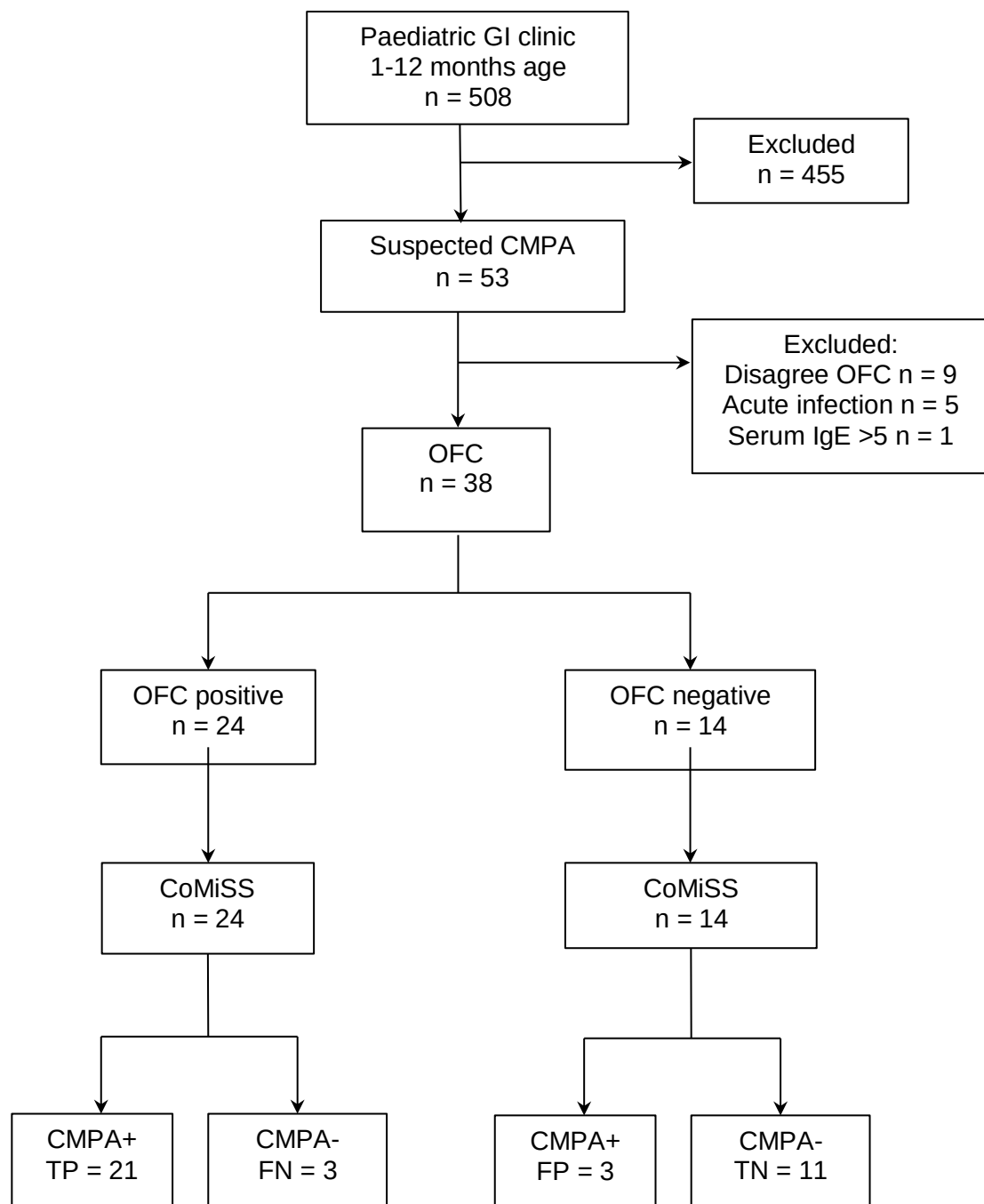
DOMAIN 4: FLOW AND TIMING	
A. Risk of Bias	
Describe any patients who did not receive the index test(s) and/or reference standard who were excluded from the 2x2 table (refer to flow diagram): All 85 infants who had oral food challenge were included in the analysis N=32 excluded after inclusion as parents refused oral food challenge Describe the time interval and any interventions between index test(s) and reference standard: Unclear when CoMiSS was completed after oral food challenge	
Was there an appropriate interval between index test(s) and reference standard? An appropriate interval is defined as up to six weeks.	Yes/no/unclear
Did patients receive the same reference standard?	Yes/no/unclear
Were all patients included in the analysis? N=31 patients' data is not included (those in which OFC was not performed)	Yes/no/unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW/HIGH/UNCLEAR

130

131

Study: Zeng 2019

Phase 2: Flow diagram.



141 **Phase 3: Risk of bias and applicability judgments.**

142

DOMAIN 1: PATIENT SELECTION	
A. Risk of Bias	
Describe methods of patient selection: All infants 1-12 months consulted at the paediatric GI clinic were selected and those with appropriate symptoms were included	
Was a consecutive or random sample of patients enrolled? All eligible infants consulted at clinic were enrolled	Yes/no/unclear
Was a case-control design avoided?	Yes/no/unclear
Did the study avoid inappropriate exclusions? Reference standard was applied to all infants that had CoMiSS score. Small number of infants were excluded because parents refused OFC or they had an infection	Yes/no/unclear
Could the selection of patients have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting): Appropriate setting in a paediatric GI clinic and score used for detecting CMPA No prior testing in the children for CMPA	
Is there concern that the included patients do not match the review question?	CONCERN: LOW/HIGH/UNCLEAR

143

144

DOMAIN 2: INDEX TEST(S)	
If more than one index test was used, please complete for each test.	
A. Risk of Bias	
Describe the index test and how it was conducted and interpreted: CoMiSS symptom score assessed Completed by GI physicians who had appropriate training on how to complete the score	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes/no/unclear
If a threshold was used, was it pre-specified? The cut-off and values for sensitivity and specificity were calculated using the data	Yes/no/unclear
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differs from the review question?	CONCERN: LOW/HIGH/UNCLEAR

145

146

DOMAIN 3: REFERENCE STANDARD	
A. Risk of Bias	
Describe the reference standard and how it was conducted and interpreted: Oral food challenge was used, the gold standard test for CMPA Conducted by GI physician using appropriate protocol. Not clear whether the physician was blinded to results of CoMiSS	
Is the reference standard likely to correctly classify the target condition? Not DBPCFC	Yes/no/unclear

Were the reference standard results interpreted without knowledge of the results of the index test?	Yes/no/ unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW/HIGH/UNCLEAR
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW/HIGH/UNCLEAR

147

148

DOMAIN 4: FLOW AND TIMING	
A. Risk of Bias	
Describe any patients who did not receive the index test(s) and/or reference standard who were excluded from the 2x2 table (refer to flow diagram): All infants who had CoMiSS had an OFC and were all included in the analysis Exclusions: refused OFC (n=9), acute infection (n=5), serum IgE > 5.0 (n=1) Describe the time interval and any interventions between index test(s) and reference standard: Not clear when OFC was done in relation to completion of CoMiSS	
Was there an appropriate interval between index test(s) and reference standard? Not clear when OFC took place	Yes/no/ unclear
Did patients receive the same reference standard?	Yes /no/ unclear
Were all patients included in the analysis?	Yes /no/ unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW/HIGH/UNCLEAR

149

Supplementary File 7. Additional test accuracy data reported by, or calculated based on data reported in, diagnostic test accuracy studies on CMPA questionnaire and CoMiSS™. Additional data consists of area under the curve (AUC) and positive and negative likelihood ratios.

Study	Cut-off score	Total infants in analysis	TP	FP	FN	TN	Sensitivity [95% CI]	Specificity [95% CI]	Reported area under the curve (AUC)	Likelihood ratios
<i>CMPA questionnaire</i>										
Gibbons, 2012	≥ 6 (max 24)	84	38	12	5	29	0.88 [0.75, 0.96]	0.71 [0.54, 0.84]	-	LR+ 3.02 LR- 0.16
Gibbons, 2012	≥ 6 (max 15)	84	38	3	9	38	0.81 [0.67, 0.91]	0.93 [0.80, 0.98]	0.89	LR+ 11.1 LR- 0.21
<i>CoMiSS™ (†or its earlier SBS version)</i>										
<i>At presentation</i>										
Prasad, 2018†	≥ 12	83	55	8	15	5	0.79 [0.67, 0.87]	0.38 [0.14, 0.68]	0.68	LR+ 1.28 LR- 0.56
Salvatore, 2019	≥ 9	47	16	4	3	24	0.84 [0.60, 0.97]	0.86 [0.67, 0.96]	0.91	LR+ 5.16 LR- 0.68
Salvatore, 2019	≥ 12	47	7	2	12	26	0.37 [0.16, 0.62]	0.93 [0.76, 0.99]	-	LR+ 5.89 LR- 0.18

Selbuz, 2020 ^s	≥ 12.5	168	-	-	-	-	-	-	0.57	-
Sirin Kose, 2019 [†]	≥ 10	49	43	-	6	-	0.88 [0.75, 0.95]	Not estimable	-	-
Sirin Kose, 2019 [†]	≥ 12	49	34	-	15	-	0.69 [0.55, 0.82]	Not estimable	-	-
Vandenplas, 2014 [†]	≥ 12	84	58	-	26	-	0.69 [0.58, 0.79]	Not estimable	-	-
Zeng, 2019	≥ 5.5	38	21	3	3	11	0.88 [0.68, 0.97]	0.79 [0.49, 0.95]	0.89	LR+ 4.08 LR- 0.41
<i>At 1-month after ED</i>										
Sirin Kose, 2019 [†]	≥ 25% decrease after 4-week ED	49	48	-	1	-	0.98 [0.89, 1.00]	Not estimable	-	-
Sirin Kose, 2019 [†]	≥ 50% decrease after 4-week ED	49	41	-	8	-	0.84 [0.70, 0.93]	Not estimable	-	-
Vandenplas, 2014 [†]	≥ 50% decrease after 4-week ED	85	45	11	14	15	0.76 [0.63, 0.86]	0.58 [0.37, 0.77]	-	LR+ 1.80 LR- 0.41

Supplementary file 8. Forest plots for test accuracy of CoMiSS tool and CMPA questionnaire.

