

Comparison of severity scorings in oral food challenges with cow's milk

Nieminen O, Palosuo K, Kukkonen K, Mäkelä M

Helsinki Skin and Allergy Hospital, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.

Corresponding author:

Otso Nieminen

The Skin and Allergy Hospital, University of Helsinki and Helsinki University Hospital, Finland.

E-mail: otso.nieminen@helsinki.fi

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: 10.18176_jiaci.0913

Key words: Oral food challenge. Cow's milk allergy. Severity scoring. Symptom. Allergic reaction.

Palabras clave: Provocación oral con alimentos. Alergia a leche de vaca. Escala de gravedad. Síntoma. Reacción alérgica.

Cow's milk allergy (CMA) is the most common food allergy in early childhood. Its diagnosis is based on clinical history, sensitization, and oral food challenges (OFC). Different scorings have been developed to objectively assess the OFC symptoms and their severity. However, the great number of scorings reflects the difficulty of fully standardizing symptom assessment [1]. We compared two widely used methods of symptom assessment, *i.e.* Hourihane's [2] and Sampson's [3] scorings, and investigated their differences in open OFCs with cow's milk. Furthermore, we compared these scorings to the Food Allergy Severity Score (FASS) [4], the only validated scoring system developed.

We challenged 135 Finnish children (median age: 1,8 years, interquartile range: 1,3 – 3,7 years) with cow's milk at the Helsinki University Skin and Allergy Hospital as previously described [5] and evaluated the reactions according to PRACTALL consensus criteria [6]. Symptom severity was assessed retrospectively by ON with Hourihane's severity scoring (HSS) [2], Sampson's severity scoring (SSS) [3], modified Hourihane's severity scoring (mHSS) [7], and modified Sampson's severity scoring (mSSS) (sFile S1). Additionally, FASS was applied to all the reactions using R studio and the code provided by the developers of FASS. [4] All symptoms evaluated by PRACTALL [6] and their respective severity scorings appear in sTable S2. ON acquired a second opinion from KP in cases of uncertainty. The level of agreement between the scorings was determined by linearly weighted Cohen's kappa with SPSS Inc. (Chicago, IL, USA) version 27 software. Sampson's grades were classified as mild (grades 1

and 2), moderate (grade 3), and severe (grades 4 and 5). Tables and graphs were prepared with Microsoft Excel (version 2211).

One-hundred-and-three (76%) of the 135 challenges were positive. Six of the positive reactions were excluded from further analysis due to inconclusive symptoms. The reaction severity classification by SSS [3], HSS [2], mSSS, mHSS [7] and oFASS-3 [4] appears in Figure 1.

The agreement on reaction severity between SSS [3] and HSS [2] was weak (weighted Cohen's kappa 0,496). Unlike SSS [3], HSS [2] was affected by the cumulative reactive dose. Abdominal pain was recognised only by HSS [2], while inspiratory stridor was not identified by SSS [3]. SSS [3] identified many symptoms not recognised by HSS [2]. These were flushing, oral allergy syndrome (OAS), lip swelling, nausea, diarrhea, loss of bowel control, swallowing difficulty, upper airway symptoms (sneezing, rhinorrhea, nasal congestion), cardiovascular symptoms (tachycardia, hypotension, severe bradycardia, dysrhythmia, cardiac arrest), dyspnea, cyanosis, feeling of pending doom, light headedness, and change in activity level. Symptoms not recognised by either scoring included persistent cough, itching and rubbing of eyes and nose, and drop in arterial blood saturation.

Some symptoms were scored differently. Vomiting was graded as moderate by HSS [2], but as mild (1 episode of vomiting) or moderate (>1 episode) by SSS [3]. Angioedema was moderate with HSS but mild with SSS, whereas generalized urticaria was either mild or moderate by HSS depending on the cumulative reactive dose, but received mild grading with SSS. [2], [3]

In our cohort, 20 of the 81 reactions scored as mild by SSS [3] were scored as moderate by HSS [2]. This was due to vomiting in 10 cases, generalized urticaria elicited by a small dose in 6 cases, and abdominal pain in 4 cases (instead graded by OAS in 2, nausea in 1, and nasal congestion in 1 case with SSS [3]). Seven cases of rhinorrhea were moderate according to SSS [3] but mild with HSS [2], as local skin reactions were used for grading. One case of dyspnea scored as severe by SSS [3] was moderate with HSS [2] due to abdominal pain. Eight cases of persistent cough were not recognised by either scoring. Instead, they were graded by vomiting in 2, local skin reactions in 3, and generalized urticaria in 1 case. The remaining 2 cases were scored by rhinorrhea with SSS [3] and local skin reactions with HSS [2].

The differences between SSS [3] and HSS [2] were especially pronounced in gastrointestinal symptoms. Vomiting was graded differently, and OAS was not recognised by HSS [2]. In

addition, HSS [2] always graded abdominal pain as moderate whereas SSS [3] neglected it altogether. Relatively common symptoms like persistent cough and itching of eyes were neglected by both scorings. HSS [2] that was originally developed for peanut challenges, does not recognise upper airway symptoms, dysrhythmias, and changes in activity level. In addition, identification of severe respiratory symptoms and laryngeal symptoms depends largely on individual interpretation in HSS [2].

Our mSSS (sFile S1) added mild and severe abdominal pain as recognised symptoms (grades 2 and 3, respectively) and graded intermittent and constant sneezing and rhinorrhea as mild (grade 2) and moderate (grade 3) symptoms [8]. Itching of eyes was added as a mild symptom. [9] Persistent cough and inspiratory stridor were added as severe grade 4 symptoms. Agreement between SSS [3] and mSSS was moderate (weighted Cohen's kappa 0,663). Comparison between HSS [2] and mHSS [7] appears in sFile S3. Their mutual agreement was weak (weighted Cohen's kappa 0,570).

FASS [4] recognises symptoms according to PRATCALL [6] consensus criteria. It uses the organs and systems affected rather than symptoms when grading reaction severity. Grade 1 (OAS) is considered mild, grades 2 and 3 moderate (1 and >1 of skin, nose/eye, digestive tract, or uterus affected, respectively), and grades 4 (larynx or bronchi) and 5 (cardiovascular or nervous system) severe. Mild and moderate reactions assessed by SSS [3] or HSS [2] were moderate in FASS [4], apart from 12 cases considered severe. These were one case of weakness, one case of dyspnea (severe in SSS [3]), and 10 cases of frequent coughing not recognised by either scoring. All cases assessed as severe by either SSS [3] or HSS [2] were also severe in FASS [4].

Using oFASS-3 [4], no reaction was mild since there were no patients with only itching of mouth or throat. 76 reactions (78%) were moderate and 21 (22%) were severe (Figure 1). Using oFASS-5 [4], fifty-four (71%) of the moderate reactions were grade 2 and 22 (29%) were grade 3. Of the severe reactions, 17 (81%) were grade 4 and 4 (19%) were grade 5.

Unlike SSS [3] and HSS [2], FASS [4] recognised both itching and rubbing of eyes and nose and frequent coughing. Compared to SSS [3], FASS [4] recognised abdominal pain and inspiratory stridor. It did not recognise nasal congestion, dyspnea, respiratory arrest, swallowing difficulty, loss of bowel control, and feeling of pending doom. Also, dysrhythmia, severe bradycardia, and cyanosis were not specified. Weighted Cohen's kappa was 0,150.

Unlike HSS [2], FASS [4] recognised flushing, lip edema, OAS, nausea, upper respiratory symptoms (rhinorrhea, sneezing), changes in activity level, weakness, tachycardia, hypotension, and diarrhea. It also better specified laryngeal symptoms, collapse, and severe respiratory symptoms. Weighted Cohen's kappa was 0,155. These kappa values are probably explained by the absence of patients with OAS as their only symptom, recognition of frequent coughing as a symptom, and the use of organs and systems to grade severity instead of symptoms. When comparing oFASS-5 [4] to Sampson's [3] grades 1 to 5, kappa value increased to 0,261 which further supports the role of different views on OAS and coughing as explaining factors for the small kappa values.

Our results evaluating reaction severity in 97 positive oral cow's milk challenges show that different severity scorings lead to different gradings. Clinicians worldwide use varied scorings, which leads into more heterogeneity. Even identical scorings show substantial inter-observer variability for symptom assessment, with kappa for mutual agreement being 0,31 to 0,46. [10] Attempts to predict reaction severity using different markers, such as specific IgE, would benefit from a unified severity scoring. Currently, only one severity scoring has been properly validated. [4] As such there is much anticipation for the DEFASE project undertaken by World Allergy Organization to further our understanding of the severity of food allergies. [1]

Acknowledgements

We want to thank Janne Burman, M.D., for contributing to the conceptualization of this study. We also want to thank the personnel of Helsinki Skin and Allergy Hospital for carrying out the challenges and the families for their participation in the study.

Funding

This work was supported by the Helsinki University Hospital Research Funds, the Allergy Research Foundation, the Pediatric Research Foundation, University of Helsinki, and the Sigrid Jusélius Foundation.

Conflict of interest

The authors declare no conflicts of interest.

List of abbreviations:

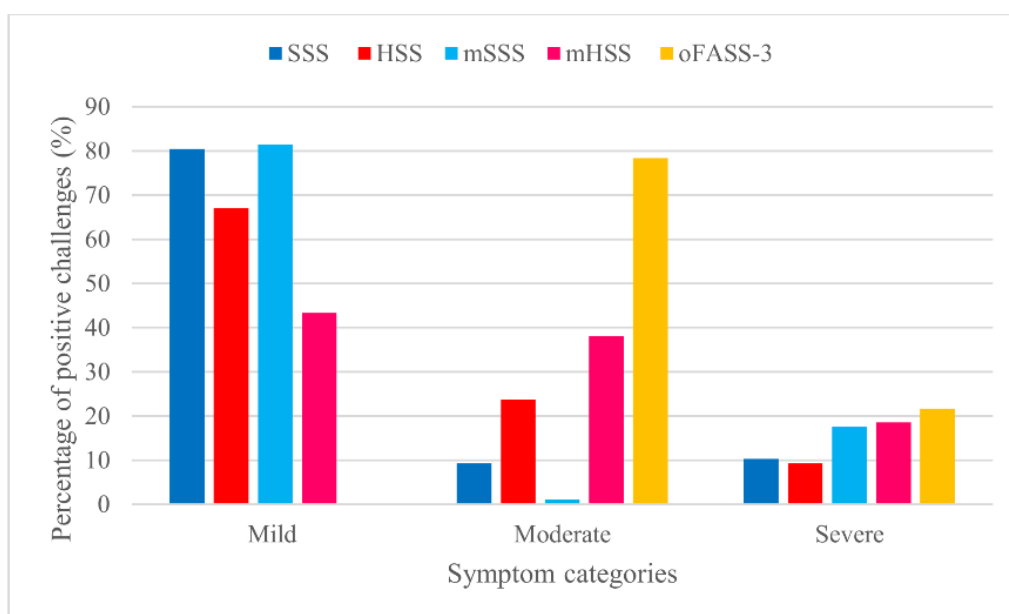
CMA (cow's milk allergy), OFC (oral food challenge), FASS (Food Allergy Severity Score) HSS (Hourihane's severity scoring), SSS (Sampson's severity scoring), mHSS (modified Hourihane's severity scoring), mSSS (modified Sampson's severity scoring), OAS (oral allergy syndrome)

References

1. Arasi S, Nurmatov U, Dunn-Galvin A, Daher S, Roberts G, Turner PJ, et al. Consensus on DEfinition of Food Allergy SEverity (DEFASE) an integrated mixed methods systematic review. *World Allergy Organ J.* 2021;14(3):100503.
2. Hourihane JOB, Grimshaw KEC, Lewis SA, Briggs RA, Trewin JB, King RM, et al. Does severity of low-dose, double-blind, placebo-controlled food challenges reflect severity of allergic reactions to peanut in the community? *Clin Exp Allergy.* 2005;35(9):1227–33.
3. Sampson HA. Anaphylaxis and Emergency Treatment. *Pediatrics.* 2003;111(6 Pt 3):1601–8.
4. Fernández-Rivas M, Gómez García I, Gonzalo-Fernández A, Fuentes Ferrer M, Dölle-Bierke S, Marco-Martín G, et al. Development and validation of the food allergy severity score. *Allergy.* 2022;77(5):1545-58.
5. Nieminen O, Palosuo K, Kukkonen K, Makela M. Molecular allergy diagnostics in predicting oral cow's milk challenge outcome in Finnish children. *Allergy Asthma Proc.* 2022;44(1):71-7.
6. Sampson HA, Gerth Van Wijk R, Bindslev-Jensen C, Sicherer S, Teuber SS, Burks AW, et al. Standardizing double-blind, placebo-controlled oral food challenges: American Academy of Allergy, Asthma & Immunology–European Academy of Allergy and Clinical Immunology PRACTALL consensus report. *J Allergy Clin Immunol.* 2012;130(6):1260–74.
7. Kukkonen AK, Pelkonen AS, Mäkinen-Kiljunen S, Voutilainen H, Mäkelä MJ. Ara h 2 and Ara 6 are the best predictors of severe peanut allergy: a double-blind placebo-controlled study. *Allergy.* 2015;70(10):1239–45.
8. Itazawa T, Adachi Y, Takahashi Y, Miura K, Uehara Y, Kameda M, et al. The severity of reaction after food challenges depends on the indication: A prospective multicenter study. *Pediatr Allergy Immunol.* 2020;31(2):167–74.
9. Muraro A, Roberts G, Clark A, Eigenmann PA, Halken S, Lack G, et al. The management of anaphylaxis in childhood: position paper of the European academy of allergology and clinical immunology. *Allergy.* 2007;62(8):857–71.
10. Van Erp FC, Knulst AC, Meijer Y, Gabriele C, Van Der Ent CK. Standardized food challenges are subject to variability in interpretation of clinical symptoms. *Clin Transl Allergy.* 2014;4(1):1–6.

Figure legends

Figure. Bar chart represents oral food challenge reaction severities assessed by the five different scorings (n=97). Reaction severities graded similarly by both SSS and HSS appear bolded in the table. In parenthesis are the scores leading to the corresponding severity grades. SSS = Sampson's severity scoring [3], HSS = Hourihane's severity scoring [2], mSSS = modified Sampson's severity scoring (sTable S1), mHSS = modified Hourihane's severity scoring [7], oFASS-3 [4].



		SSS		
		Mild (1-2)	Moderate (3)	Severe (4-5)
HSS	Mild (1-5)	58	7	0
	Moderate (6-13)	20	2	1
	Severe (14-23)	0	0	9