Wheat challenge in self-reported gluten sensitivity: A comparison of scoring methods

G.I. Skodje[1], C. Henriksen[2], T. Salte[3], T. Drivenes[3], I. Toleikyte[1], A.M. Lovik[1], M.B. Veierød[2,4], K.E.A. Lundin[5,6,7]

[1]Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital Rikshospitalet, 0424 Oslo, Norway; E-mails: g.i.skodje@medisin.uio.no, ievtol@ous-hf.no, astridmargretel@gmail.com

[2]Department of Nutrition, Institute of Basic Medical Sciences, University of Oslo; E-mail: christine.henriksen@medisin.no.no

[3]Faculty of Medicine, University of Oslo, 0317 Oslo, Norway; E-mails: trude_salte@hotmail.com, thea_drivenes@hotmail.com

[4]Oslo Centre for Biostatistics and Epidemiology, Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, 0317 Oslo, Norway; E -mail: m.b.veierod@medisin.uio.no

[5]Department of Gastroenterology, Oslo University Hospital Rikshospitalet, 0424 Oslo, Norway; E –mail: k.e.a.lundin@medisin.uio.no

[6]Centre for Immune Regulation, University of Oslo, 0424 Oslo, Norway

[7]Institute of Clinical Medicine, University of Oslo, 0318 Oslo, Norway

Corresponding author: G. I. Skodje, e-mail: g.i.skodje@medisin.uio.no

Abstract

Background: The condition non-coeliac gluten sensitivity (NCGS) is clinically similar to coeliac disease, but lack objective diagnostic criteria. Symptom relief on gluten-free diet followed by gluten containing food challenge may confirm the condition in clinical settings.

Aim: To describe the results of an open bread challenge in patients with suspected NCGS, and to compare the results with recently suggested cut-offs for symptom change.

Material and methods: Fifty-six patients (12 males) self-instituted on gluten-free diet with negative coeliac disease diagnostics were examined for NCGS by an open bread challenge. Symptoms were reported by Gastrointestinal Symptom Rating Scale, IBS-version (GSRS-IBS) and visual analogue scale (VAS). Results were retrospectively compared to the Salerno and Monash cut-offs for symptom change.

Results: Forty-seven patients were diagnosed with NCGS. Total GSRS-IBS score and overall symptoms by VAS increased significantly in NCGS ($p<0.001$), but not in non-NCGS patients ($p<0.12$ and $p=0.08$, respectively). Total GSRS-IBS challenge score and overall symptoms by VAS were significantly higher in NCGS than in non-NCGS patients (53 vs 37, $p=0.004$ and 76 vs 39 mm, $p=0.02$, respectively). Applying the Salerno and Monash cut-offs, 63 and 75% would be classified with NCGS, respectively. According to total GSRS–IBS absolute agreement was lowest between clinician's diagnosis and Salerno cut-off (63%) and highest between Salerno and Monash cut-offs (88%).

Conclusion: Clinician diagnosed 85% with NCGS. The proportion of NCGS was lower according to the Salerno and Monash cut-offs. The Salerno cut-off should be the starting point for a common definition of symptom change.

43 Keywords: Non-coeliac gluten sensitivity, gluten challenge, diagnostics, symptom score,

44 Salerno Experts' Criteria

Introduction

Non-coeliac gluten sensitivity (NCGS) is a relatively new entity of gluten related disorders that differs from coeliac disease (CD) and wheat allergy (WA), but without a precise and accepted definition [1]. NCGS is so far understood as a condition associated with the experience of various intestinal and extra-intestinal symptoms in response to ingestion of wheat, rye and barley, followed by resolution of symptoms on dietary removal of these foods, when CD and WA is properly excluded [2]. NCGS was first described in the early 1980's in the United Kingdom [3], and in recent years there has been an increase in reports of the condition . There are no reliable biomarkers for the diagnosis. The true prevalence is therefore still unknown, also because many patients are self-diagnosed and start a gluten-free diet (GFD) without a proper examination [4]. NCGS may also overlap with conditions such as irritable bowel syndrome (IBS) and food hypersensitivities that also lack diagnostic precision [5].

Currently diagnosis of NCGS is based on exclusion of other diseases, combined with patient's experience of symptom relief on GFD [6]. Elimination of dietary gluten followed by a standardized gluten challenge may confirm the condition, as improvement of symptoms after gluten exclusion is regarded as a diagnostic criterion for NCGS. Open food challenges has been widely used for clinical purposes, especially in dietician guided assessment of various food hypersensitivities, also suggested to facilitate the diagnose of NCSG [7]. However, there has not been any standardization of the gluten containing vehicle, how to record symptoms, or how to define symptom change. Thus, in 2009 Oslo University Hospital standardized an open bread challenge to assess NCGS in a clinical setting. The classification of NCGS or non-NCGS has been performed by a gastroenterologist (hereafter called clinician), based on clinical interview, symptom scores by VAS and validated questionnaires, but without any predefined standard for symptom change.

Recently the Salerno Expert group recommended blinded placebo controlled gluten challenge for the confirmation of NCGS, though, without stating any best suited gluten containing vehicle. A threshold of >30 % symptom change was suggested to define a patient responsive to challenge (Salerno cut-off) [8]. Another threshold of change, minimum 20 mm on 100 mm visual analogue scale (VAS), has recently been used in a controlled research setting by Australian researchers (Monash cut-off) [9] .

Our first aim was to describe the results of the open bread challenge in patients with suspected NCGS in Norway 2009-2015. Further, these data provide a unique opportunity to retrospectively compare the results of our clinical evaluation with the recently suggested Salerno and Monash cut-offs for symptom change.

Material and methods

Dietician's and clinician's reports were analyzed from medical records of patients investigated for NCGS from January 2009 to September 2015 at the Department of Gastroenterology and the Division of Clinical Nutrition, Oslo University Hospital, Rikshospitalet, composing a retrospective quality control study.

All patients were investigated according to a protocol prepared prior to inclusion period. Referrals came from local hospital or general practitioner. The patients were self-instituted on gluten-free diet followed by self-reported symptom relief. Coeliac disease was considered to be ruled out if the patient had had a normal duodenal biopsy while on a gluten-containing diet, or if they were negative for the coeliac disease compatible genotypes HLA-DQ2.5 or –DQ8. Patients with clinically defined wheat allergy were deemed not suitable for challenge. An experienced dietitian carried out nutrition assessment including evaluation of dietary intake, history of food related symptoms and diet adherence by a standardized interview. Height and weight were measured, body mass index (kg/m^2) calculated and instructions for the open bread challenge were given.

Open bread challenge

Patients underwent a 3-14 days open challenge of four slices of white bread (120 g) considered to equate eight grams of gluten per day. The length of the period changed over the years: The majority (n=41) were challenged for 3 days, but two patients were challenged for five days, two for seven days (n=2) and 11 for 14 days. The longer period was due to patients reporting that three days of challenge would not be enough to catch the symptoms. The bread was purely wheat based and retrospective analysis showed that 100 grams of white bread contained 6.8 g of gluten (SD 5.2, 8.4) [10] and 2.1 g of fructans (analyzed by the Monash University). Patients were told otherwise to keep their diet

consistent with baseline throughout the challenge. However, consistency was not checked. Gastrointestinal symptoms were recorded by a Norwegian translation of the Gastrointestinal Symptom Rating Scale, irritable bowel syndrome-version (GSRS-IBS), reflecting three days of their normal situation before the challenge (baseline), and after the challenge. GSRS-IBS is a self-administered 13-items questionnaire, with a seven-point Likert scale for each item (ranging from 1=’no symptom’ to 7=’severe symptom’) [11]. The questionnaire has previously been used to monitor response during gluten challenge in coeliac and NCGS patients [12]. The five sub-dimension scores (pain, bloating, constipation, diarrhea and satiety), were calculated retrospectively and were not available for the clinician’s classification. Abdominal pain, bloating and overall symptoms were recorded daily on a 100 mm VAS. The dietician calculated the symptom scores. The clinician evaluated the challenge response of each patient, without any predefined standard for symptom change, and concluded with positive or negative diagnosis for NCGS. Reasons for not getting the diagnosis was 1) negative or small change of symptoms in response to challenge or 2) high baseline symptom score indicating lack of symptom control despite living on gluten-free diet. The diagnosis was used by the patients when they applied for reimbursement for gluten-free food (as defined by National Administration of Labour and Welfare).

Comparison of scoring methods

Results of the clinical evaluation of NCGS were compared with the Salerno and Monash criteria. In 2015 the Salerno criteria was suggested to standardize the investigation of NCGS for clinical and research purposes [8]: A symptom increase of ≥30 % when exposed to a given amount of gluten in a standardized challenge was considered significant worsening. Since symptom recording in the current study also included 100 mm VAS, the Monash criteria was also used for comparison: A symptom increase of ≥20 mm on VAS was considered a clinically relevant worsening of symptoms when challenged to standardized gluten amount in a research setting [9]. To make the scoring methods comparable, GSRS-IBS scores ≥30 and ≥20 % of total were classified as NCGS with Salerno and Monash criteria, respectively. Change in score on visual analogue scale ≥30 mm was classified as NCGS with the Salerno criteria.

For simplicity, the comparison of the prevalence of NCGS between the scoring methods focuses on the total GSRS-IBS score.

Statistics

Results are presented as frequencies (%), means (standard deviations, SDs) or medians (interquartile range, IQR). Patients were classified as NCGS or non-NCGS. Changes from baseline to challenge for the whole group were tested by the Wilcoxon signed rank test. The two groups were compared by the Mann Whitney U test. Comparison between the two groups as regards sex was performed by a chi-square test. Spearman correlation coefficient (r) was calculated to estimate the associations between the symptom scores for overall symptoms, abdominal pain and bloating by VAS and total score for GSRS-IBS and the pain and bloating dimensions.

Agreement between the scoring methods was estimated by absolute agreement (the proportion of patients on the diagonal, P_A) and Kappa (κ). Strength of agreement was evaluated according to the Kappa cut-offs given by Landis & Koch [13]: $=0.01 \leq \kappa \leq 0.20$ poor, $0.21 \leq \kappa \leq 0.40$ fair, $0.41 \leq \kappa \leq 0.60$ moderate, $0.61 \leq \kappa \leq 0.80$ substantial and $0.81 \leq \kappa \leq 0.99$ excellent agreement. McNemar test was used to test symmetry. Specific agreement was also calculated, expressing separately the agreement for the diagnosis of NCGS (positive agreement, P_{pos}) and non-NCGS (negative agreement, P_{neg}) using the formulas in de Vet et al [14]. Two-sided p-values <0.05 were considered significant. Statistical analysis was carried out using the statistical software IBM SPSS.22.

Retrospective power analysis was performed. Mean (SD) mm overall symptoms assessed by VAS during challenge was 67 (26) for NCGS (n=36) and 41 (15) for non-NCGS (n=6), and significantly different (p=0.001). With a significance level of 0.05, we had 88 % power to detect this difference in VAS score between NCGS and non-NCGS patients with group sizes of 36 and 6, respectively (calculated in Stata).

The manuscript was prepared according to the standard criteria in the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement (http://www.strobe-statement.org).

Ethics

The study was approved by the local Privacy Commissioner for Research at Oslo University Hospital, Division for patient security and quality the 22nd of April 2015, registered with the project identification code 2014/16821 - Gluten challenge in suspected non-coeliac gluten sensitivity - a retrospective study. Written informed consent was obtained post-investigation from all patients.

Results

Study sample

Of 63 patients referred for NCGS investigation from January 2009 to September 2015, 56 (12 males) completed the open bread challenge protocol (5 did not return symptom records and 2 had a diagnosis of inflammatory bowel disease). Diet interview showed that all patients adhered well to the gluten-free diet at baseline.

All 56 patients completed the GSRS-IBS questionnaire and 42 completed the VAS. The evaluation by the consultant clinician resulted in 47 patients (85 %) receiving a NCGS diagnosis and 9 not receiving the diagnosis, hereafter referred to as NCGS and non-NCGS patients, respectively. No significant differences were found between the groups as regards sex, BMI and duration of GFD ($0.50 \leq p \leq 0.95$, Table 1), but the non-NCGS patients were significantly older than the NCGS patients (means 54 and 38 years, respectively, p=0.005).

Table 1 Characteristics of the study sample

	Total (n=56)	NCGS (n=47)	Non-NCGS (n=9)	p-value[a]
Sex n (m/f)	12/44	10/37	2/7	0.95
Age (years), mean (SD)	41 (15.4)	38 (14.7)	54 (11.6)	0.005
BMI (kg/m^2), median (IQR)	22.6 (20, 27)	22.9 (20, 27)	21.9 (21, 25)	0.50
GFD (months), median (IQR)	16 (4, 37)	18 (9, 44)	12 (1, 36)	0.65

NCGS=Non coeliac gluten sensitivity, SD=Standard deviation, BMI=Body mass index
IQR=interquartile range. GFD=gluten free diet.
[a]Chi-square test for gender and Mann Whitney U test for the other variables

Assessment of NCGS by clinician

Symptom scoring by total GSRS-IBS, GSRS dimensions, and VAS for overall symptoms, abdominal pain and bloating increased significantly from baseline to challenge in the total sample ($p<0.001$, Table S1). Figure 1 shows that the NCGS patients had significant increase in gastrointestinal symptoms as response to the challenge ($p<0.001$), whereas the non-NCGS patients did not change significantly in symptom score from baseline to challenge ($0.12\leq p\leq 0.67$). Total GSRS-IBS score at baseline was significantly higher in non-NCGS than in NCGS patients ($p=0.005$, Table S2). Total GSRS-IBS score during challenge and GSRS-IBS score for pain, bloating, diarrhea dimensions were significantly higher in the NCGS compared to non-NCGS patients ($0.004\leq p\leq 0.03$), while no significant difference was observed for constipation and satiety dimensions ($p=0.79$ and $p=0.14$, respectively).

Table S1 Baseline and challenge symptom score assessed by GSRS-IBS (total n=56, domains n=55) and VAS (n=42)

	Baseline		Challenge		Change score		
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	p-value[a]
Total GSRS-IBS	20.3 (7.4)	18.5 (14, 23)	49.6 (14.2)	50.5 (38, 61)	29.2 (15.5)	32.0 (18, 41)	<0.001
Pain	3.5 (1.7)	3.0 (2, 4)	8.4 (2.9)	8.0 (7, 10)	4.9 (3.2)	5.0 (3, 7)	<0.001
Bloating	5.6 (2.8)	5.0 (3, 7)	14.2 (4.7)	15.0 (11, 18)	8.6 (5.7)	10.0 (5, 12)	<0.001
Constipation	3.1 (2.1)	2.0 (2, 3)	5.6 (3.9)	5.0 (2, 8)	2.7 (4.3)	2.0 (0, 5)	<0.001
Diarrhea	6.0 (3.0)	5.0 (4, 7)	14.2 (6.6)	13.0 (9, 19)	8.2 (7.6)	7.0 (3, 14)	<0.001
Satiety	2.9 (1.5)	2.0 (2, 3)	6.7 (3.5)	7.0 (3, 9)	3.9 (3.5)	4.0 (1, 7)	<0.001
VAS							
Overall symptoms	10.3 (12.3)	6.0 (2, 18)	58.9 (25.8)	62.0 (45, 85)	48.6 (29.3)	47.4 (30, 75)	<0.001
Abdominal pain	9.4 (13.0)	5.5 (1, 15)	58.3 (26.4)	62.0 (36, 80)	48.9 (28.2)	46.3 (35, 77)	<0.001
Bloating	13.1 (14.0)	7.0 (2, 14)	63.2 (26.5)	63.2 (32, 81)	50.1 (29.0)	53.5 (27, 75)	<0.001

GSRS=gastrointestinal symptom rating scale- IBS version, missing in domains n=1, VAS=visual analogue scale 0-100 mm missing n=14

SD=standard deviation. IQR=interquartile range.

[a]Wilcoxon signed rank test

185

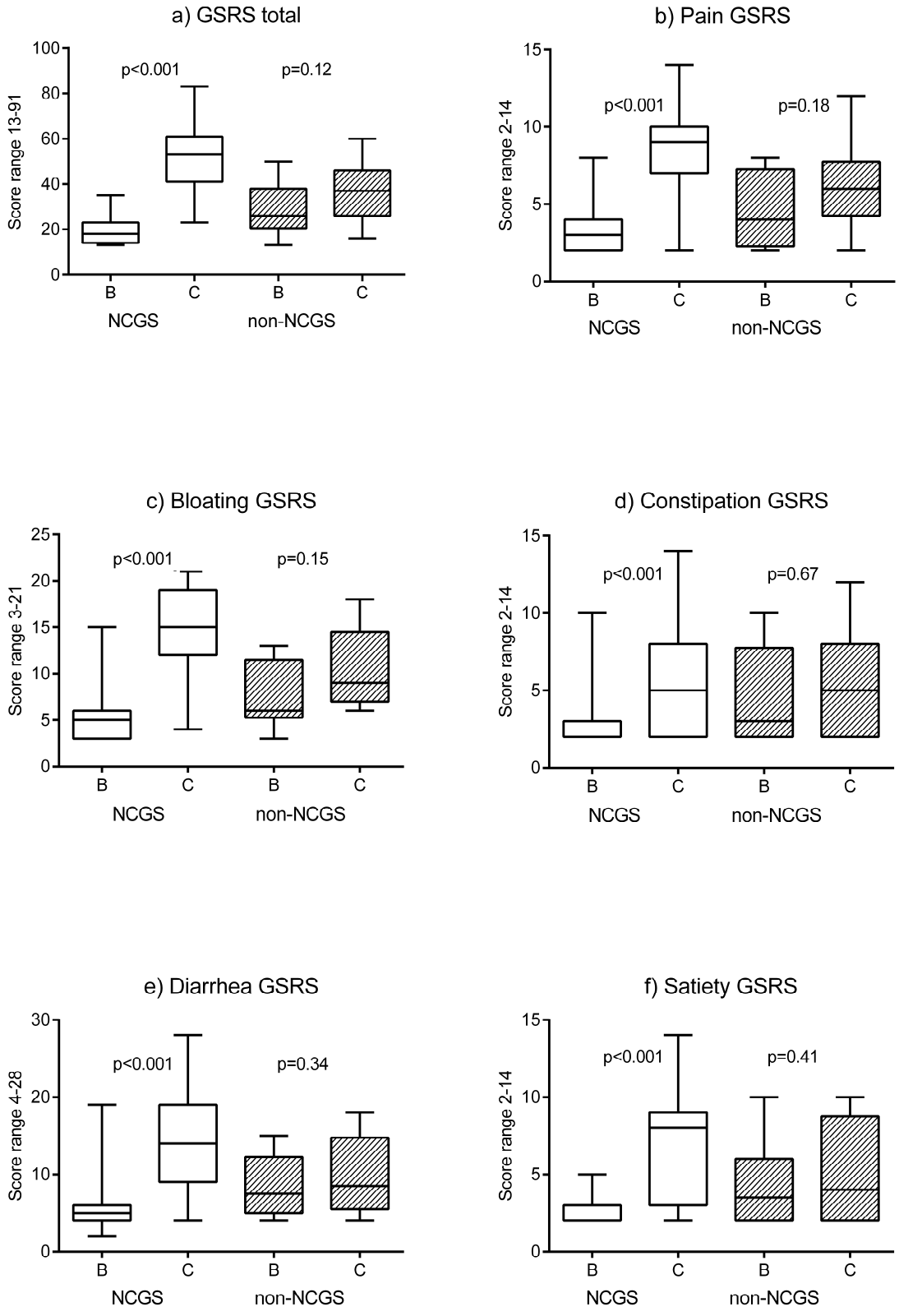

Figure 1 Box plot of baseline (B) and challenge (C) scores for total gastrointestinal symptom rating scale-(GSRS)-IBS and GSRS-dimensions in non-coeliac gluten sensitivity (NCGS) (n=47) and non-NCGS (n=8). P-values from the Mann Whitney U test.

Table S2 Baseline and challenge symptom score by GSRS-IBS and VAS in NCGS (n=47) and non-NCGS (n=8)

	Median baseline score (IQR)			Median challenge score (IQR)			Median change score (IQR)		
	NCGS (n=47)	Non-NCGS (n=8)	p-value[a]	NCGS (n=47)	Non-NCGS (n=8)	p-value[a]	NCGS (n=47)	Non-NCGS (n=8)	p-value[a]
GSRS-IBS total	18 (14,23)	26 (21,38)	0.005	53 (41,61)	37 (26,46)	0.004	36 (27,44)	8 (-7,16)	<0.001
Pain	3 (2,4)	4 (2,7)	0.19	9 (7,10)	6 (4,8)	0.02	5 (4,7)	1 (0,5)	0.003
Bloating	5 (3,6)	6 (5,12)	0.06	15 (12,19)	9 (7,15)	0.01	10 (7,13)	3 (-2,5)	0.001
Constipation	2 (2,3)	3 (2,8)	0.22	5 (2,8)	5 (2,8)	0.79	3 (0,5)	1 (-4,6)	0.31
Diarrhea	5 (4,6)	8 (5,12)	0.05	14 (9,19)	9 (6,15)	0.03	9 (5,14)	2 (0,5)	0.001
Satiety	2 (2,3)	4 (2,6)	0.08	8 (3,9)	4 (2,9)	0.14	5 (1,7)	0 (0,1)	0.001
VAS	(n=36)	(n=6)		(n=36)	(n=6)		(n=36)	(n=6)	
Overall symptoms	6 (2,16)	20 (9,40)	0.07	76 (54,85)	39 (30,53)	0.02	55 (38, 77)	19 (-2, 34)	0.001
Abdominal pain	6 (1,12)	21 (7,41)	0.02	70 (43,82)	28 (24,39)	0.004	50 (37, 78)	10 (14, 28)	<0.001
Bloating	6 (1,13)	7 (4,23)	0.30	69 (34,82)	33 (24,53)	0.03	55 (29, 76)	28 (11, 39)	0.03

GSRS-IBS=gastrointestinal symptom rating scale – IBS version, missing in domains non-NCGS n=1, VAS=visual analogue scale 0-100 mm, NCGS=non-coeliac gluten sensitivity, IQR=inter quartile range, SD=standard deviation

[a]Mann Whitney U test

192 The NCGS patients had significant increase in symptom score by VAS from baseline to challenge
193 ($p<0.001$, Table 2). The non-NCGS patients showed some increase in bloating and overall symptoms,
194 though not significant $p=0.05$ and $p=0.08$, respectively. All VAS scores during challenge were
195 significantly higher in NCGS than in the non-NCGS patients ($p\leq0.03$, Table S2). The change in
196 symptom score from baseline to challenge was significantly higher in NCGS patients for overall
197 symptoms, abdominal pain, and bloating ($p\leq0.02$)

Table 2 Baseline and challenge symptom score[a] assessed by 100 mm visual analogue scale in NCGS (n=36) and non-NCGS (n=6)

	NCGS			Non-NCGS			Change score		
	Baseline	Challenge	p-value[b]	Baseline	Challenge	p-value[b]	NCGS	Non-NCGS	p-value[b]
Overall symptoms (mm)	6 (2, 16)	76 (54, 85)	<0.001	20 (9, 40)	39 (30, 53)	0.08	55 (38, 77)	19 (-2, 34)	0.001
Abdominal pain (mm)	6 (1, 12)	70 (43, 82)	<0.001	21 (7, 41)	28 (24, 39)	0.60	50 (37, 78)	10 (14, 28)	<0.001
Bloating (mm)	6 (1, 13)	69 (34, 82)	<0.001	7 (4, 23)	33 (24, 53)	0.05	55 (29, 76)	28 (11, 39)	0.03

NCGS=Non-coeliac gluten sensitivity

[a]Median (interquartile range)

[b]Mann Whitney U test

198

Correlation coefficients stratified by group are presented in Table S3. Significantly positive correlation was found between the GSRS-IBS and VAS scores for overall symptoms, abdominal pain, and bloating at baseline and challenge for NCGS ($0.49 \leq r \leq 0.69$, $p \leq 0.01$).

Table S3 Spearman's correlation coefficient between scores of GSRS-IBS and VAS for abdominal pain, bloating and overall symptoms (n=42)

	Baseline				Challenge				Change score			
GSRS vs VAS	NCGS (n=36)	p-value	Non NCGS (n=6)	p-value	NCGS (n=36)	p-value	Non NCGS (n=6)	p-value	NCGS (n=36)	p-value	Non NCGS (n=6)	p-value
Overall symptoms	0.53	0.001	0.83	0.04	0.70	<0.001	0.99	< 0.001	0.56	0.001	0.99	0.001
Abdominal pain	0.69	0.01	0.98	<0.001	0.61	<0.001	0.40	0.51	0.56	0.001	0.82	0.09
Bloating	0.49	0.003	0.46	0.43	0.61	<0.001	0.46	0.43	0.48	0.003	0.96	0.005

202

Comparison with Salerno and Monash criteria

In the retrospective analysis, according to total GSRS-IBS score NCGS would be classified in 63 % by the Salerno cut-off and 75 % by the Monash cut-off, as compared to the 85 % diagnosed by the clinician (Table 3). Classification as NCGS according to overall symptoms by VAS score would be 74 % by Salerno cut-off and 81% by the Monash cut -off, as compared to the 86 % diagnosed by the clinician.

Absolute agreement (P_A) according to total GSRS-IBS between the clinician's diagnosis and the new classification according to the Salerno cut-off was 75 % (κ =0.40), increased to P_A= 85 % between clinician and Monash (κ =0.51) and was highest between the Salerno and Monash classification P_A=88 % (κ=0.71, Table 3). Absolute agreement according to overall symptoms by VAS was higher between the clinician's diagnosis and the Salerno cut-off classification, P_A=88 % (κ=0.60), than between clinician diagnosis and Monash cut-off classification, P_A=86 % (κ=0.42). For overall symptoms assessed by VAS, applying the Salerno and Monash cut-offs showed highest absolute agreement, P_A=93 % (κ=0.76).

Table 3 Prevalence (P_0) and absolute agreement (P_A) between clinician-, Salerno- and Monash classifications, for GSRS-IBS and VAS

		Prevalence (P_0)			Clinician-Salerno		Clinician-Monash		Salerno-Monash	
		Clinician	Salerno	Monash	P_A	Kappa	P_A	Kappa	P_A	Kappa
	N	n (%)	n (%)	n (%)	%		%		%	
GSRS-IBS total	56	47 (85)[b]	35 (63)	42 (75)	75	0.40	85	0.51	88	0.71
Pain	55[a]		34 (62)	45 (82)	69	0.26	85	0.47	80	0.53
Bloating			38 (69)	42 (76)	80	0.45	84	0.48	93	0.82
Constipation			17 (31)	27 (49)	34	0.03	53	0.07	82	0.63
Diarrhea			24 (44)	34 (62)	62	0.23	73	0.35	82	0.65
Satiety			26 (47)	31 (56)	62	0.27	67	0.28	91	0.82
VAS	42[a]	36 (86)[b]								
Overall symptoms			31 (74)	34 (81)	79	0.35	81	0.32	93	0.80
Abdominal pain			33 (79)	36 (86)	88	0.60	86	0.42	93	0.76
Bloating			30 (71)	37 (88)	71	0.18	76	0.06	83	0.51

GSRS-IBS=gastrointestinal symptom rating scale – IBS version, VAS=visual analogue scale 0-100 mm

[a] Total n, missing GSRS domains n=1, VAS n=14

[b] Clinician evaluated only for total score

Table 4 shows disagreement. The distribution was skewed towards more patients diagnosed with NCGS by clinician than would have been if the Salerno cut-off had been applied. There was a lack of symmetry for all GSRS-IBS dimensions in the comparison between clinician's diagnosis and Salerno cut-off classification ($p \leq 0.01$). Disagreement between clinician's diagnosis and Monash cut-off classification was lower but in the same direction. Here, there was no lack of symmetry according to total GSRS-IBS ($p=0.18$). Disagreement between Salerno and Monash was skewed oppositely towards more patients classified with NCGS when applying the Monash cut-off than the Salerno cut-off. Significant lack of symmetry was found according to total GSRS-IBS ($p=0.02$). Disagreement according to overall symptoms by VAS was lower than for total GSRS-IBS in all comparisons, and there was no lack of symmetry ($0.25 \leq p \leq 1.0$).

Table 4 Disagreement (S_D) and symmetry (Sym) between clinician-, Salerno- and Monash classifications, for GSRS-IBS and VAS

		Clinician-Salerno			Clinician-Monash			Salerno-Monash		
		Disagreement (S_D)			Disagreement (SD)			Disagreement (SD)		
		$NCGS_{Clin}$/ non-$NCGS_{Sal}$	Non-$NCGS_{Clin}$/ $NCGS_{Sal}$	Sym p-value[a]	$NCGS_{Clin}$/ non-$NCGS_{Mon}$	Non-$NCGS_{Clin}$/ $NCGS_{Mon}$	Sym p-value[a]	$NCGS_{Sal}$/ non-$NCGS_{Mon}$	Non-$NCGS_{Sal}$/ $NCGS_{Mon}$	Sym p-value[a]
	N	n	n		n	n		n	n	
GSRS-IBS total	56	13	1	0.002	7	2	0.18	0	7	0.02
Pain	55[b]	15	2	0.002	5	3	0.73	0	11	0.001
Bloating		10	1	0.01	7	2	0.18	0	4	0.13
Constipation		32	2	<0.001	23	3	<0.001	0	10	0.002
Diarrhea		23	0	<0.001	14	1	0.001	0	10	0.002
Satiety		21	0	<0.001	17	1	<0.001	0	5	0.06
VAS	42[b]									
Overall symptoms		7	2	0.18	5	3	0.73	0	3	0.25
Abdominal pain		4	1	0.38	3	3	1.00	0	3	0.25
Bloating		9	3	0.15	4	5	1.00	0	7	0.02

GSRS-IBS=gastrointestinal symptom rating scale – IBS version, VAS=visual analogue scale 0-100 mm, NCGS=Non coeliac gluten sensitivity

[a] McNemar

[b] Total n, missing GSRS domains n=1, VAS n=14

231 Table 5 shows specific agreement expressing agreement on NCGS as positive agreement (P_{pos}) and
232 agreement on non-NCGS as negative agreement (P_{neg}). According to total GSRS-IBS there was higher
233 positive than negative agreement in all comparisons. Lowest specific agreement was found between
234 clinician's diagnosis and Salerno classification (P_{pos}=83 %, P_{neg}=53 %) and highest between Salerno
235 and Monash classification (P_{pos}=91 %, P_{neg}=80 %).

236

237

Table 5 Specific positive (P_{pos}) and negative (P_{neg}) agreement [a] between clinician-, Salerno- and Monash classifications, for GSRS-IBS and VAS

		Clinician-Salerno		Clinician-Monash		Salerno-Monash	
		P_{pos}	P_{neg}	P_{pos}	P_{neg}	P_{pos}	P_{neg}
	N	%	%	%	%	%	%
GSRS-IBS total	56	83[e]	53	90	61	91	80
Pain	55 [b]	79	41	91	56	86	65
Bloating		87	56	90	57	95	87
Constipation		48	26	65	28	77	85
Diarrhea		68	41	81	48	83	81
Satiety		71	43	77	44	91	91
VAS	42 [b]						
Overall symptoms		87	47	89	43	95	84
Abdominal pain		93	67	92	50	96	80
Bloating		82	33	88	18	90	59

GSRS-IBS=gastrointestinal symptom rating scale – IBS version, VAS=visual analogue scale 0-100 mm, NCGS=Non coeliac gluten sensitivity

[a] Calculated according to the formulas provided in de Vet et al[13]

[b] Total n, missing GSRS domains n=1, VAS n=14

238

Discussion

Open bread challenges were performed in a clinical setting and resulted in 85 % classified by the clinician as having NCGS. These 47 patients had significantly higher increase in symptom score from baseline to challenge, and higher score during challenge than the non-NCGS patients. A lower number of patients would have been classified with NCGS if applying the recent Salerno and Monash cut-offs. Absolute agreement was lowest between clinician's diagnosis and Salerno cut-off, and highest between Salerno and Monash cut-offs.

Previous studies assessing NCGS

NCGS is described as a diagnosis of elimination in individuals self-reporting symptom relief on GFD [2,4,15], but clinical trials with DBPC challenges are increasing [16,17]. Volta et al studied patients from 38 medical centers over one year and found 486 patients with suspected NCGS. The presumption of NCGS was based on the medical record, and effect of GFD was estimated by a retrospective self-administered questionnaire [18]. The presumption of NCGS was strengthened in 80 % (391 patients) which is close to the 85 % of the present study. However, the authors point out that a standardized challenge had not been performed to confirm the condition.

Zanini et al investigated 35 individuals and classified 31 % as NCGS based on the individual's ability to identify the gluten containing flour in a DBPC challenge [17]. Here, individuals that did not score on symptoms during challenge, but correctly identified gluten, could falsely be classified as having NCGS. Symptom score should be the primary effect measure of a challenge. However, the false positives may also occur in the situation of the clinician's evaluation without any predefined cut-off for symptom change, as in the present study.

Raju et al who confirmed the diagnosis in 82 % of patients with suspected NCGS, after a 6 week open bread challenge in a clinical setting [19]. Apart from a longer challenge, this report has a setting and method comparable to the present study.

Finally, Elli at al was the first to follow the Salerno Experts' criteria in their randomized clinical trial of DBPC gluten challenge where 14 % was classified as having NCGS [20]. A nocebo effect of the open challenge may explain why the present study still had higher proportion of NCGS classified by the Salerno cut-off, 63 %, as compared to Elli et al. The patient selections also differed in that Elli et al excluded all patients already following a gluten-free diet, whereas all patients in the present study had been gluten-free for an average of 16 months.

Comparison with Salerno and Monash cut-offs

Absolute agreement of P_A=75 % between clinician's diagnosis and Salerno cut-off indicated only fair agreement according to the Kappa of 0.40 (Table 3), but the probability of clinician's diagnosis and Salerno cut-off to agree on NCGS based on total GSRS-IBS score was P_{Pos}=83 %, as expressed by the positive agreement (Table 5). However, the probability of the two to agree on non-NCGS was only P_{neg}=53 %. This is as result of the skewed distribution illustrated by the disagreement of 13 of 14 patients classified as non-NCGS where clinician diagnosed NCGS (Table 4). The Monash cut-off was less strict than the Salerno cut-off. The absolute agreement was therefore higher between clinician's diagnosis and Monash cut-off (P_A=85 %), though only moderate according the Kappa of 0.51 (Table 3). Here, the probability of agreement on NCGS was high, P_{Pos}=90 % (Table 5), and is explained by a symmetric and lower degree of disagreement (Table 4). Since Salerno and Monash cut-offs were similar scoring methods that differed only by 10 % or 10 mm in cut-off, the highest absolute agreement was seen here, P_A=88 % (Table 3). Kappa value of 0.71 indicated substantial agreement. Both positive and negative agreement was high indicating a probability of agreement on NCGS of P_{pos}=91 % and on non-NCGS of P_{neg}=80 % (Table 5). However, given a lower cut-off of Monash, more patients were classified with NCGS by Monash than by Salerno cut-off, as shown by the lack of symmetry in the disagreement, again according to total GSRS-IBS score (Table 4). Consequently, of the three scoring methods, the Salerno cut-off generated the lowest proportion of NCGS.

Patients in the present study were already on a gluten-free diet with perceived symptom relief. If the classification by Salerno cut-off is closest to the true prevalence, patients would be falsely classified with NCGS by clinician and Monash cut-off. The National Administration of Labour and Welfare in

Norway requires a confirmation of the diagnosis each third year. There is no requirement of a repeated challenge, but there would certainly be necessary to re-evaluate the condition in patients of the present study, preferably with a blinded placebo controlled procedure.

Appraisal of methods

The strength of the study was that NCGS investigation was confirmed by a standardised open bread challenge, not only based on elimination criteria. The high correlation between symptom scoring by GSRS-IBS and VAS indicated internal consistency and reliable reporting by the patients. .
DBPCFC is stated as the gold standard to investigate adverse reactions to food. However, the method is poorly standardized [21], resource demanding and therefore suited for controlled research rather than clinical settings, where it has several pitfalls [22]. In contrast to DBPCFC, the open white bread challenge mimics real life and is more likely to reproduce the reported reactions, where the double blinded procedure risk to generate false negatives. Open food challenges are still accepted in clinical settings [7,23] and are the first step diagnostic tool in exploring adverse reactions to food when patients are self-instituted on an elimination diet, and may be followed by blinded challenge where there is a need for confirmation [24].

The non-NCGS patients reported significantly higher symptoms at baseline than the NCGS patients, indicating lack of symptom control on gluten-free diet, and therefore correctly classified with negative outcome. However, both groups had increased symptoms during challenge. We expect the nocebo effect to be present in both groups, still the non-NCGS patients showed less increase in symptoms than the NCGS patients. The higher baseline symptom score in the non-NCGS patients could possibly be explained by dietary fermentable carbohydrates (FODMAPs) as was the case in the study of Biesiekierski et al. [9]. In that case, a low FODMAP diet would have been a better suited dietary treatment [25].

The open challenge is also a weakness of the study a as the method is known to overestimate positive outcomes caused by the nocebo effect [26]. Further, wheat contains fermentable carbohydrates like fructans, known to reduce symptoms in individuals with IBS if removed from the diet [25]. The

316 content of fructans in white bread must be taken into account since they coexist with gluten. The
317 present study cannot distinguish between reactions caused by gluten or fructans. Last, the clinician's
318 evaluation lacked standardization, especially in terms of definition of symptom change. This may have
319 resulted in overestimation of positive outcomes when compared to the Salerno and Monash cut-offs.
320 Lack of diagnostic criteria and defined cut-offs will easily lead to overestimation because the clinician
321 is dependent on the subjective reports of the patients. The Salerno Expert's Criteria are therefore
322 welcome as a common diagnostic method for this condition.

323 The comparison of the clinician's diagnosis with the Salerno and Monash cut-offs is the first attempt
324 to compare definitions of symptom change in diagnostics of NCGS. However, none of the scoring
325 methods represent the gold standard. Their ability to reflect the truth is uncertain and the Salerno
326 experts state that the ≥30 % increment needs validation [8]. The Monash 20 mm cut-off is of unknown
327 origin, but to compare, 20 mm has been defined as "little more" and 40 mm as "much more" of a
328 symptom [27], and 22 mm as minimal clinical significant change for measuring nausea in adults [28].
329 The Monash cut-off was included in the comparison given that it was used in a DBPC gluten
330 challenge in self-reported gluten sensitive individuals, which was a setting similar to what was
331 suggested by the Salerno Experts' Criteria. In addition to GSRS the Salerno criteria include a ten point
332 numeric rating scale for the recording of symptoms, which is not very different from a 100 mm VAS.
333 The Monash cut-off therefore strengthened the comparison.

334 Higher agreement and lower disagreement was observed for the VAS variables in all comparisons,
335 which may indicate that VAS is a more universal tool for measuring symptoms in NCGS than the
336 GSRS-IBS questionnaire.

337 Specific agreement (P_{pos}, P_{neg}) was chosen because it is suggested as the most appropriate measure for
338 inter-observer variation for clinicians [14]. When the question concerns the reliability of a test, the
339 ability to distinguish between NCGS and non-NCGS reliability measures as Cohen's Kappa are
340 preferred. Here, the interest was also the observer agreement, such as knowing if the clinician's NCGS
341 diagnosis would be in agreement with the NCGS diagnosis classified by the Salerno and Monash cut-

off's, in other words the precision. The clinician's interest in this observer variation calls for an absolute measure of agreement. The result of a skewed distribution such as in the present study is a low relative measure (low Kappa). However, when clinicians ask whether patients with NCGS can be distinguished from non-NCGS, they have the individual patient in mind and are not considering the distribution. They would want to know the probability of agreement, the so-called specific agreement.

The finding of absolute agreement between clinician's diagnosis and the Salerno cut-off of 75 % and value for Kappa 0.40 indicated low or moderate agreement [13]. Though, the precision was higher, given by a positive agreement of 83 %. However, regardless of good precision, both classifications may have low accuracy, since Salerno cut-off's ability to actually diagnose NCGS is uncertain. There is a chance that they all have missed the bull's eye.

Conclusion

This open bread challenge resulted in 85 % diagnosed as NCGS. The prevalence was lower according to the Salerno and Monash cut-offs classifications. Best agreement was found between the Salerno and Monash cut-offs, and lowest between clinician's diagnosis and Salerno cut-off. There is a probability of overestimation in the clinician's diagnosis of methodological reasons. However, it is not possible to tell which scoring method that is closest to the truth, as there is no real gold standard to compare with.

Scientists and clinicians should use a standardized diagnostic method for diagnosing NCGS and a common definition of symptom change. Independently of the present results, a good starting point is to use the Salerno cut-off. Future studies need to test thethe performance of the Salerno Expert's Criteria.

Acknowledgments

Thanks to all the patients that gave written consent to use information from their patient record.

Author Contributions

GIS, MBV and KEAL designed the work. GIS, TS, TD, IT and AL collected the patient record information and systemized the data. GIS performed the statistical analysis, drafted the manuscript and was responsible for the manuscript revisions; MBV contributed to the statistical analysis and interpretation of data; CH, MBV and KEAL contributed to the interpretation of the results and the writing of the manuscript. All authors critically revised the manuscript and approved the final version.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Ludvigsson, J.F.; Leffler, D.A.; Bai, J.C.; Biagi, F.; Fasano, A.; Green, P.H.; Hadjivassiliou, M.; Kaukinen, K.; Kelly, C.P.; Leonard, J.N.*, et al.* The oslo definitions for coeliac disease and related terms. *Gut* **2013**, *62*, 43-52.
2. Lundin, K.E.; Alaedini, A. Non-celiac gluten sensitivity. *Gastrointestinal endoscopy clinics of North America* **2012**, *22*, 723-734.
3. Cooper, B.T.; Holmes, G.K.; Ferguson, R.; Thompson, R.A.; Allan, R.N.; Cooke, W.T. Gluten-sensitive diarrhea without evidence of celiac disease. *Gastroenterology* **1980**, *79*, 801-806.
4. Catassi, C.; Bai, J.; Bonaz, B.; Bouma, G.; Calabrò, A.; Carroccio, A.; Castillejo, G.; Ciacci, C.; Cristofori, F.; Dolinsek, J.*, et al.* Non-celiac gluten sensitivity: The new frontier of gluten related disorders. *Nutrients* **2013**, *5*, 3839-3853.
5. Verdu, E.F.; Armstrong, D.; Murray, J.A. Between celiac disease and irritable bowel syndrome: The "no man's land" of gluten sensitivity. *The American journal of gastroenterology* **2009**, *104*, 1587-1594.
6. Ludvigsson, J.F.; Bai, J.C.; Biagi, F.; Card, T.R.; Ciacci, C.; Ciclitira, P.J.; Green, P.H.; Hadjivassiliou, M.; Holdoway, A.; van Heel, D.A.*, et al.* Diagnosis and management of adult coeliac disease: Guidelines from the british society of gastroenterology. *Gut* **2014**, *63*, 1210-1228.
7. Vazquez-Roque, M.; Oxentenko, A.S. Nonceliac gluten sensitivity. *Mayo Clin. Proc.* **2015**, *90*, 1272-1277.
8. Catassi, C., Elli, L., Bonaz, B., Bouma, G., Carroccio, A., Castillejo, G., Cellier, C., Cristofori, F., de Magistris, L., Dolinsek, J., Dietrich, W., Francavilla, R., . How the diagnosis of non-celiac gluten sensitivity (ncgs) should be confirmed: The salerno experts' criteria. *Nutrients* **2015**.
9. Biesiekierski, J.R.; Peters, S.L.; Newnham, E.D.; Rosella, O.; Muir, J.G.; Gibson, P.R. No effects of gluten in patients with self-reported non-celiac gluten sensitivity after dietary reduction of fermentable, poorly absorbed, short-chain carbohydrates. *Gastroenterology* **2013**, *145*, 320-328 e321-323.
10. Lovik, A.; Skodje, G.; Bratlie, J.; Brottveit, M.; Lundin, K.E. Diet adherence and gluten exposure in coeliac disease and self-reported non-coeliac gluten sensitivity. *Clin. Nutr.* **2015**. *145*, 320-328 e321-323.
11. Wiklund, I.K.; Fullerton, S.; Hawkey, C.J.; Jones, R.H.; Longstreth, G.F.; Mayer, E.A.; Peacock, R.A.; Wilson, I.K.; Naesdal, J. An irritable bowel syndrome-specific symptom questionnaire: Development and validation. *Scand.J.Gastroenterol.* **2003**, *38*, 947-954.
12. Brottveit, M.; Vandvik, P.O.; Wojniusz, S.; Lovik, A.; Lundin, K.E.; Boye, B. Absence of somatization in non-coeliac gluten sensitivity. *Scandinavian journal of gastroenterology* **2012**, *47*, 770-777.
13. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159-174.
14. de Vet, H.C.; Mokkink, L.B.; Terwee, C.B.; Hoekstra, O.S.; Knol, D.L. Clinicians are right not to like cohen's kappa. *BMJ* **2013**, *346*, f2125.
15. Sapone, A.; Bai, J.C.; Ciacci, C.; Dolinsek, J.; Green, P.H.; Hadjivassiliou, M.; Kaukinen, K.; Rostami, K.; Sanders, D.S.; Schumann, M.*, et al.* Spectrum of gluten-related disorders: Consensus on new nomenclature and classification. *BMC Med.* **2012**, *10*, 13.
16. Carroccio, A.; Mansueto, P.; Iacono, G.; Soresi, M.; D'Alcamo, A.; Cavataio, F.; Brusca, I.; Florena, A.M.; Ambrosiano, G.; Seidita, A.*, et al.* Non-celiac wheat sensitivity diagnosed by double-blind placebo-controlled challenge: Exploring a new clinical entity. *The American journal of gastroenterology* **2012**, *107*, 1898-1906; quiz 1907.
17. Zanini, B.; Baschè, R.; Ferraresi, A.; Ricci, C.; Lanzarotto, F.; Marullo, M.; Villanacci, V.; Hidalgo, A.; Lanzini, A. Randomised clinical study: Gluten challenge induces symptom recurrence in

only a minority of patients who meet clinical criteria for non-coeliac gluten sensitivity. *Alimentary pharmacology & therapeutics* **2015**, *42*, 968-976.
18. Volta, U.; Bardella, M.T.; Calabro, A.; Troncone, R.; Corazza, G.R. An italian prospective multicenter survey on patients suspected of having non-celiac gluten sensitivity. *BMC Med.* **2014**, *12*, 85.
19. Raju, S.A.; Mooney, P.D.; Aziz, I.; Kurien, M.; Sanders, D.S. Letter: Gluten challenge in the era of noncoeliac gluten sensitivity – a change in clinical practice? *Alimentary pharmacology & therapeutics* **2016**, *43*, 656-656.
20. Elli, L.; Tomba, C.; Branchi, F.; Roncoroni, L.; Lombardo, V.; Bardella, M.T.; Ferretti, F.; Conte, D.; Valiante, F.; Fini, L.*, et al.* Evidence for the presence of non-celiac gluten sensitivity in patients with functional gastrointestinal symptoms: Results from a multicenter randomized double-blind placebo-controlled gluten challenge. *Nutrients* **2016**, *8*. 10.3390/nu8020084.
21. Asero, R.; Fernandez-Rivas, M.; Knulst, A.C.; Bruijnzeel-Koomen, C.A. Double-blind, placebo-controlled food challenge in adults in everyday clinical practice: A reappraisal of their limitations and real indications. *Curr. Opin. Allergy Clin. Immunol.* **2009**, *9*, 379-385.
22. Niggemann, B.; Beyer, K. Pitfalls in double-blind, placebo-controlled oral food challenges. *Allergy* **2007**, *62*, 729-732.
23. Boyce, J.A.; Assa'ad, A.; Burks, A.W.; Jones, S.M.; Sampson, H.A.; Wood, R.A.; Plaut, M.; Cooper, S.F.; Fenton, M.J.; Arshad, S.H.*, et al.* Guidelines for the diagnosis and management of food allergy in the united states: Report of the niaid-sponsored expert panel. *J. Allergy Clin. Immunol.* **2010**, *126*, S1-58.
24. Yao, C.K.; Gibson, P.R.; Shepherd, S.J. Design of clinical trials evaluating dietary interventions in patients with functional gastrointestinal disorders. *The American journal of gastroenterology* **2013**, *108*, 748-758.
25. Halmos, E.P.; Power, V.A.; Shepherd, S.J.; Gibson, P.R.; Muir, J.G. A diet low in fodmaps reduces symptoms of irritable bowel syndrome. *Gastroenterology* **2014**, *146*, 67-75 e65.
26. Venter, C.; Pereira, B.; Grundy, J.; Clayton, C.B.; Roberts, G.; Higgins, B.; Dean, T. Incidence of parentally reported and clinically diagnosed food hypersensitivity in the first year of life. *J. Allergy Clin. Immunol.* **2006**, *117*, 1118-1124.
27. Bulloch, B.; Tenenbein, M. Assessment of clinically significant changes in acute pain in children. *Acad. Emerg. Med.* **2002**, *9*, 199-202.
28. Meek, R.; Kelly, A.M.; Hu, X.F. Use of the visual analog scale to rate and monitor severity of nausea in the emergency department. *Acad. Emerg. Med.* **2009**, *16*, 1304-1310.